

CC transfected with the recombinant expression construct, screening a
CC compound for modulating human hwt1 gene activity in cells expressing the
CC human hwt1 gene product, a cell membrane preparation (or cytosolic
CC preparation or nuclear preparation) comprising a human hwt1 gene product
CC (or derivative) and identifying a compound that induces or increases hwt1
CC gene expression in mammalian cells. Culturing bone marrow or peripheral
CC blood culture in the presence of compounds identified by the methods is
CC useful for increasing retention of primitive CD34⁺ haematopoietic stem
CC cells. Alternatively the method comprises introducing a retroviral or
CC lentiviral expression construct into the cells. The compounds are useful
CC for haematopoietic stem cell development and maintenance and for
CC inhibiting leukaemia cell growth and treating cancer. The present
CC sequence encodes Hwt1.
CC
CC
XX

SQ Sequence 2328 BP; 716 A; 496 C; 520 G; 596 T; 0 U; 0 Other;
Query Match 100.0%; Score 2328; DB 9; Length 2328;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGATCTTGGGTGAACAAGAGGAAAGCTTAGACATGTTAAAGATCAAAAACAGGT	60
Db	1	ATGATCTTGGGTGAACAAGAGGAAAGCTTAGACATGTTAAAGATCAAAAACAGGT	60
QY	61	TCTTCAGGCATTAATNGTAGGTTAAGCACTAACCAATTCGCGCTGACATCCGCTCCGAG	120
Db	61	TCTTCAGGCATTAATNGTAGGTTAAGCACTAACCAATTCGCGCTGACATCCGCTCCGAG	120
QY	121	TGGGCTTAATACAGATACATTGACATATAACCACTGATGGAAAGCCGAAGACTCCGT	180
Db	121	TGGGCTTAATACAGATACATTGACATATAACCACTGATGGAAAGCCGAAGACTCCGT	180
QY	181	TCAGCTCTCTTTTCAACGAAAGATCTAATTTGGAAAGTGTCAGCTTTTGAATGGAACG	240
Db	181	TCAGCTCTCTTTTCAACGAAAGATCTAATTTGGAAAGTGTCAGCTTTTGAATGGAACG	240
QY	241	ATATTAATTTTAACTTAAGAAAGACTACAGCAAAAGGTTACTGAAGTTTATGTAAGCCGG	300
Db	241	ATATTAATTTTAACTTAAGAAAGACTACAGCAAAAGGTTACTGAAGTTTATGTAAGCCGG	300
QY	301	AATGAGAGGATGTGAGATTAACGATCACTTTAACAATTGAACCTTCCACCTACATACCA	360
Db	301	AATGAGAGGATGTGAGATTAACGATCACTTTAACAATTGAACCTTCCACCTACATACCA	360
QY	361	ACTGTTTGACGTTATATATATATATTTTTCAGAGGCTTTGAAATATATGAAATTTGCA	420
Db	361	ACTGTTTGACGTTATATATATATATTTTTCAGAGGCTTTGAAATATATGAAATTTGCA	420
QY	421	CAAAATTGACGAATTAATTAATTAACCCAAATGACCAATTAATTAATTCAGATCAAGGTTG	480
Db	421	CAAAATTGACGAATTAATTAATTAACCCAAATGACCAATTAATTAATTCAGATCAAGGTTG	480
QY	481	GTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGATAGAAAGACAGATCATGCTGCG	540
Db	481	GTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGATAGAAAGACAGATCATGCTGCG	540
QY	541	ACTGACGTTAGCCATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTCAATGTTCACTTT	600
Db	541	ACTGACGTTAGCCATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTCAATGTTCACTTT	600
QY	601	TATCATCAGACAGAGAAATATAATTTCAAGAACAGATTTCCAAAGAACTAATATAGGTTTA	660
Db	601	TATCATCAGACAGAGAAATATAATTTCAAGAACAGATTTCCAAAGAACTAATATAGGTTTA	660
QY	661	GTTGTTCTTAACAATTAATTAACAATAAGACATACAGATGGAATATATGACATGGGACAG	720
Db	661	GTTGTTCTTAACAATTAATTAACAATAAGACATACAGATGGAATATATGACATGGGACAG	720
QY	721	AATCCCAAGAGCACTTTTAAAGAACCGACGCTGGAAGTCAAGTTCCTTAAGATACTAC	780
Db	721	AATCCCAAGAGCACTTTTAAAGAACCGACGCTGGAAGTCAAGTTCCTTAAGATACTAC	780
QY	781	AGGAAGCAATACACCAAGAGTACCCAGCTTGGAAGACGCTGTCTTGGTCAAGCCAGCCC	840

D	b	781	AGAGAGCAATACAAACCAAGAGATCAACCGACTTGAGAGAGCTGCTGTTGTCAGGCACACC	840
O	y	841	AAGAGAAAGCGGGGCCCTGGGGGGGACATGCGCAGGGCTGCGCATGCTCATTCCTGAGACTG	900
D	b	841	AAGGAAAGGGGGGCCCTGGGGGGGACATGCGCAGGGCTGCGCATGCTCATTCCTGAGACTG	900
O	y	901	TGCAATCTTACAGGCTCTAACTGATTTAAATCGTATGATTTTAACTGTAGTAAGACTTA	960
D	b	901	TGCAATCTTACAGGCTCTAACTGATTTAAATCGTATGATTTTAACTGTAGTAAGACTTA	960
O	y	961	GCGGTTCATCAAGACTTAATCTTCAAGCAAAAGGCAAGCTGTGAAGTGGGACCACTATTGAT	1020
D	b	961	GCGGTTCATCAAGACTTAATCTTCAAGCAAAAGGCAAGCTGTGAAGTGGGACCACTATTGAT	1020
O	y	1021	TACATTCATTAATAACGATTAATGTTCAAGAGGAGCTTGACACTGGGGTTTGAGCTTTGAT	1080
D	b	1021	TACATTCATTAATAACGATTAATGTTCAAGAGGAGCTTGACACTGGGGTTTGAGCTTTGAT	1080
O	y	1081	TCCAACTTACTGTCTTCTCAGAGAGATTTTGCAACAGAAAAGATTCAACCAAGGTGGA	1140
D	b	1081	TCCAACTTACTGTCTTCTCAGAGAGATTTTGCAACAGAAAAGATTCAACCAAGGTGGA	1140
O	y	1141	AAAACTTATATACATTCACAAATTGGCAGATTGTCACAAAGAAACAAGGTGCACCA	1200
D	b	1141	AAAACTTATATACATTCACAAATTGGCAGATTGTCACAAAGAAACAAGGTGCACCA	1200
O	y	1201	TTAATTAGTGTAAAGCCATAGATTAATCTGGCTGTGATCTATACGCGAAGAAATTATGAA	1260
D	b	1201	TTAATTAGTGTAAAGCCATAGATTAATCTGGCTGTGATCTATACGCGAAGAAATTATGAA	1260
O	y	1261	GCAGCCAAATTCATTCGATACAAAATCTATTTTAAATGTCACACGCGCAAGGGCATGCAATG	1320
D	b	1261	GCAGCCAAATTCATTCGATACAAAATCTATTTTAAATGTCACACGCGCAAGGGCATGCAATG	1320
O	y	1321	AGAAAGCAATTAATGATTTGAAAGTGGATGACAGAACTGAAGACTTATTAAGTCTTAAG	1380
D	b	1321	AGAAAGCAATTAATGATTTGAAAGTGGATGACAGAACTGAAGACTTATTAAGTCTTAAG	1380
O	y	1381	CAAAAGCTCAGCAGACACCCAGATGTTGTCTGTCTGTTGTCTAACTAATCCGGAAGAAC	1440
D	b	1381	CAAAAGCTCAGCAGACACCCAGATGTTGTGTCTGTCTGTTGTCTAACTAATCCGGAAGAAC	1440
O	y	1441	AAATACGATGCTAATTAATAAAATTAACCTGTGTACAGATTGACCTTAAAGACTTATGATG	1500
D	b	1441	AAATACGATGCTAATTAATAAAATTAACCTGTGTGTACAGATTGACCTTAAAGACTTATGATG	1500
O	y	1501	GTGGCCCGGAACCTTAGGCAAAACAGCAAACTGTCACTGGCCATTGCTACAAATTTGCCCTA	1560
D	b	1501	GTGGCCCGGAACCTTAGGCAAAACAGCAAACTGTCACTGGCCATTGCTACAAATTTGCCCTA	1560
O	y	1561	CAGATGAACTGCAAGATGGGAGGAGAGACTCTGAGGGGTGAACATCCCTGTAAGTCTGTG	1620
D	b	1561	CAGATGAACTGCAAGATGGGAGGAGAGACTCTGAGGGGTGAACATCCCTGTAAGTCTGTG	1620
O	y	1621	ATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCGGAGTCAATTGCGAGAA	1680
D	b	1621	ATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCGGAGTCAATTGCGAGAA	1680
O	y	1681	TTTGTGTCAGACATCAATGAAAGGATGACCCGCTGGTCTTCAACGCTGATATTTCAAGAT	1740
D	b	1681	TTTGTGTCAGACATCAATGAAAGGATGACCCGCTGGTCTTCAACGCTGATATTTCAAGAT	1740
O	y	1741	AGAGGACAGAGAGCTGTATGATGGGCTCAAACTCTGCTGCAAGCGCTCTGAGGGCTTGG	1800
D	b	1741	AGAGGACAGAGAGCTGTATGATGGGCTCAAACTCTGCTGCAAGCGCTCTGAGGGCTTGG	1800
O	y	1801	AATAGCTGCATATGATATATGCTCCAGCGGATCACTGTGATCCCGGATGGGATGAGAGAC	1860
D	b	1801	AATAGCTGCATATGATATATGCTCCAGCGGATCACTGTGATCCCGGATGGGATGAGAGAC	1860
O	y	1861	GGCAGGCTGAAAACACTGTGTGAACCTACAGATGGCAAGTTTTTGATTTGTCTAAATCC	1920
D	b	1861	GGCAGGCTGAAAACACTGTGTGAACCTACAGATGGCAAGTTTTTGATTTGTCTAAATCC	1920

Db	1861	GGCCAGCTGAAAAACACTGCTGTGAACATGACGAAGGCCACAGCTTTTGTGATTTGCTAAATATCC	1920
Oy	1921	ATTGGTAGAGGTTACACACCTTACACTAACCGTAAATTTGGTGTGAAGAAAAAGTGAACACC	1980
Db	1921	ATTGGTAGAGGTTACACACCTTACACTAACCGTAAATTTGGTGTGAAGAAAAAGTGAACACC	1980
Oy	1981	AGATTTTGTGCTAGCTGTGAAGGAAGCTTCAGAAATCCACTTCTCGAAGATTATTGAT	2040
Db	1981	AGATTTTGTGCTAGCTGTGAAGGAAGCTTCAGAAATCCACTTCTCGAAGATTATTGAT	2040
Oy	2041	GTAGAGGTTACCAAGCCAGAAATGATATGACTTTTATATCGTAGCCAGGCTGTGAAGAT	2100
Db	2041	GTAGAGGTTACCAAGCCAGAAATGATATGACTTTTATATCGTAGCCAGGCTGTGAAGAT	2100
Oy	2101	GGTAGTGTGTTCTCCACACATATACATGTCATCTATGACCAACAGCGGCTGAAGCCAGAC	2160
Db	2101	GGTAGTGTGTTCTCCACACATATACATGTCATCTATGACCAACAGCGGCTGAAGCCAGAC	2160
Oy	2161	CACATACAGCGGTTAGACCTTACAGGCTGTGCCACATCTATTACACTGTGGCCAGGTCAATT	2220
Db	2161	CACATACAGCGGTTAGACCTTACAGGCTGTGCCACATCTATTACACTGTGGCCAGGTCAATT	2220
Oy	2221	CGTGTTCCTGCTCCCTTGGCAGTAGAGCCCAACAGCGGCTTTCTGTGTGGCCAGGATTT	2280
Db	2221	CGTGTTCCTGCTCCCTTGGCAGTAGAGCCCAACAGCGGCTTTCTGTGTGGCCAGGATTT	2280
Oy	2281	CACAGAGAGCCCAATCTGTCACTGTCAAAACCGCTTTACTACCTCTAA	2328
Db	2281	CACAGAGAGCCCAATCTGTCACTGTCAAAACCGCTTTACTACCTCTAA	2328
RESULT 2	AAA07588		
ID	AAA07588	standard; DNA; 3472 BP.	
XX	AAA07588;		
DT	29-AUG-2000	(first entry)	
DE	Human piwi gene, designated piwi.		
XX	piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;		
KW	anaemia; immunodeficiency; male infertility; human; ds.		
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	167..2752	
FT		/tag= a	
FT		/product= "hiwi"	
FT		/transl_except= (pos:192..394; aa:Xaa)	
FT		/transl_except= (pos:1073..1075; aa:Xaa)	
FT		/transl_except= (pos:2369..2370; aa:Xaa)	
FT		/note= "Xaa= Leu or Ile"	
XX	WO200032039-A1.		
XX	08-JUN-2000.		
XX	03-DEC-1999;	99WO-US028764.	
XX	04-DEC-1998;	98US-0110901P.	
XX	(UYDU-) UNITV DUKE.		
XX	Liu H;		
XX	WPI; 2000-412085/35.		
XX	P-PSDB; AAY90235.		
XX	piwi family nucleic acids, polypeptides, and antibodies, useful in gene		
XX	therapy of diseases such as cancer and in various research and diagnostic		
XX	applications.		

XX Claim 19; Page 189-194; 201pp; English.

XX This sequence encodes the human p161 family protein, designated h1w1. The
CC piwi family nucleic acids and polypeptides are used in gene therapy of
CC diseases such as cancer and also in various research and diagnostic
CC applications. The sequences can also be used to treat tissue dystrophy,
CC anaemia, immunodeficiency, and male infertility

XX
SQ Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 0 U; 3 Other;

Query Match 99.3%; Score 2311; DB 3; Length 3472;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 214; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACACAGGACGACCTTGAACCATGTTAAAGAAATCAAAAACAGGTTCTTCAG 67
DB TTGGTGTGAATACAGGACGACCACTTACCATGTTAAAGAAATCAAAAACAGGTTCTTCAG 491
QY 68 GCATTATAGTAAAGGTTTAAAGCATTAACCATTTCCGGCTGACATCCCGTCCGAGTGGGCT 127
DB GCATTATAGTAAAGGTTTAAAGCATTAACCATTTCCGGCTGACATCCCGTCCGAGTGGGCT 551
QY 128 TATATCAGTATCACAATTGACTATTAACCCACTGATGAGAGCCAGAAAGACTCCGTTCACTC 187
DB TATATCAGTATCACAATTGACTATTAACCCACTGATGAGAGCCAGAAAGACTCCGTTCACTC 611
QY 188 TTCTTTTTCACACAGGAAGTCTAATTGGAAGTGTCATGCTTTGATGGAACATATATAT 247
DB TTCTTTTTCACACAGGAAGTCTAATTGGAAGTGTCATGCTTTGATGGAACATATATAT 671
QY 612 TTCTTTTTCACACAGGAAGTCTAATTGGAAGTGTCATGCTTTGATGGAACATATATAT 671
DB TTCTTTTTCACACAGGAAGTCTAATTGGAAGTGTCATGCTTTGATGGAACATATATAT 671
QY 248 TTTTAACTTAAAGAATAACGAAAGAGTTTACTGAAGTTTATGTAAGACCCGGAATGAG 307
DB TTTTAACTTAAAGAATAACGAAAGAGTTTACTGAAGTTTATGTAAGACCCGGAATGAG 731
QY 672 TTTTAACTTAAAGAATAACGAAAGAGTTTACTGAAGTTTATGTAAGACCCGGAATGAG 731
DB TTTTAACTTAAAGAATAACGAAAGAGTTTACTGAAGTTTATGTAAGACCCGGAATGAG 731
QY 308 AGGATGTAGAGATTAACGATCATCTTTTAAACAAATGAACTCCACTACATCAACCACTTGTT 367
DB AGGATGTAGAGATTAACGATCATCTTTTAAACAAATGAACTCCACTACATCAACCACTTGTT 791
QY 732 AGATGTAGAGATTAACGATCATCTTTTAAACAAATGAACTCCACTACATCAACCACTTGTT 791
DB AGATGTAGAGATTAACGATCATCTTTTAAACAAATGAACTCCACTACATCAACCACTTGTT 791
QY 368 TGCAGTTCTAATATATATATTTTTCAGAGGCTTTTGAATATCATGAATTTGCAAAATG 427
DB TGCAGTTCTAATATATATATTTTTCAGAGGCTTTTGAATATCATGAATTTGCAAAATG 851
QY 792 TGCAGTTCTAATATATATATTTTTCAGAGGCTTTTGAATATCATGAATTTGCAAAATG 851
DB TGCAGTTCTAATATATATATTTTTCAGAGGCTTTTGAATATCATGAATTTGCAAAATG 851
QY 428 GACGAAATTAATATTAACCAATGACCAATGATATTCGAAGTCAACAGTTGATGATTT 487
DB GACGAAATTAATATTAACCAATGACCAATGATATTCGAAGTCAACAGTTGATGATTT 911
QY 852 GACGAAATTAATATTAACCAATGACCAATGATATTCGAAGTCAACAGTTGATGATTT 911
DB GACGAAATTAATATTAACCAATGACCAATGATATTCGAAGTCAACAGTTGATGATTT 911
QY 488 GGCCTGAGCTTCACTTCCATCCCTTCAGATGAAACAGCATCATGCTTCGACATGAG 547
DB GGCCTGAGCTTCACTTCCATCCCTTCAGATGAAACAGCATCATGCTTCGACATGAG 971
QY 912 GGCCTGAGCTTCACTTCCATCCCTTCAGATGAAACAGCATCATGCTTCGACATGAG 971
DB GGCCTGAGCTTCACTTCCATCCCTTCAGATGAAACAGCATCATGCTTCGACATGAG 971
QY 548 TTAGCGATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTTCATGTTCAACTTTTATCATC 607
DB TTAGCGATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTTCATGTTCAACTTTTATCATC 103
QY 972 TTAGCGATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTTCATGTTCAACTTTTATCATC 103
DB TTAGCGATAAAGTCCCTGGAAGTGAAGCTGTTTGGATTTTCATGTTCAACTTTTATCATC 103
QY 608 AGACGAAGAAATTAATTTCAAGAACAGTTTCCAAAGAACTATATAGTTTAAAGTTGTC 667
DB AGACGAAGAAATTAATTTCAAGAACAGTTTCCAAAGAACTATATAGTTTAAAGTTGTC 109
QY 1032 AGACGAAGAAATTAATTTCAAGAACAGTTTCCAAAGAACTATATAGTTTAAAGTTGTC 109
DB AGACGAAGAAATTAATTTCAAGAACAGTTTCCAAAGAACTATATAGTTTAAAGTTGTC 109
QY 668 TTACCAAGATTAACAATAAGACATATCAAGATGATGATTTGACTGGAGCCAGATATCCA 727
DB TTACCAAGATTAACAATAAGACATATCAAGATGATGATTTGACTGGAGCCAGATATCCA 115
QY 1092 TTACCAAGATTAACAATAAGACATATCAAGATGATGATTTGACTGGAGCCAGATATCCA 115
DB TTACCAAGATTAACAATAAGACATATCAAGATGATGATTTGACTGGAGCCAGATATCCA 115
QY 728 AGAGCACTTTTAAGAAAGCCGAGCTCTGAAGTCAAGCTTCTTAATATATCAAGAAAGC 787
DB AGAGCACTTTTAAGAAAGCCGAGCTCTGAAGTCAAGCTTCTTAATATATCAAGAAAGC 121
QY 1152 AGAGCACTTTTAAGAAAGCCGAGCTCTGAAGTCAAGCTTCTTAATATATCAAGAAAGC 121
DB AGAGCACTTTTAAGAAAGCCGAGCTCTGAAGTCAAGCTTCTTAATATATCAAGAAAGC 121
QY 788 AATACAAACAAGATATCAACGACTTGAAGACAGCTCTTGATGACCAAGCCAAAGAGA 847
DB AATACAAACAAGATATCAACGACTTGAAGACAGCTCTTGATGACCAAGCCAAAGAGA 127
QY 1212 AATACAAACAAGATATCAACGACTTGAAGACAGCTCTTGATGACCAAGCCAAAGAGA 127
DB AATACAAACAAGATATCAACGACTTGAAGACAGCTCTTGATGACCAAGCCAAAGAGA 127
QY 848 GGGGGGGGCTTGGGGGACATGCGAGGGCTTGCATGCTCATCTGAGCTCTGATATC 907
DB GGGGGGGGCTTGGGGGACATGCGAGGGCTTGCATGCTCATCTGAGCTCTGATATC 907

CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner, and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotide encodes and encodes
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.

XX Sequence 2579 BP; 797 A; 542 C; 613 G; 627 T; 0 U; 0 Other;

Query Match 94.5%; Score 2200.2; DB 9; Length 2579;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGGTGAACACAGGCGAAGCTTGAACATGTTAAAGATCAAAAAGGTTCTTCAAG 67
DB 319 TTGGGTGAATACAGGCGAAGCTTGAACATGTTAAAGATCAAAAAGGTTCTTCAAG 378
QY 68 GCATTTAGTAAGGTTAAGCACTAACATTTCCGGCTGACATCCGCTCCCAAGTGGGCT 127
DB 379 GCATTTAGTAAGGTTAAGCACTAACATTTCCGGCTGACATCCGCTCCCAAGTGGGCT 438
QY 128 TATATCAGATCACTGATTAACCCACTGATGGAAGCCAGAAAGCTCCGTCAAGTC 187
DB 439 TATATCAGATCACTGATTAACCCACTGATGGAAGCCAGAAAGCTCCGTCAAGTC 498
QY 188 TTCTTTTCAACGAAAGATCTAATGGAAGGTCATGCTTTTGAAGAACAGATTAAT 247
DB 499 TTCTTTTCAACGAAAGATCTAATGGAAGGTCATGCTTTTGAAGAACAGATTAAT 558
QY 248 TTTTAACTAAAGACTACAGCAAAAGTTACTGAAGTTTGAAGAACCCGGAATGAG 307
DB 559 TTTTAACTAAAGACTACAGCAAAAGTTACTGAAGTTTGAAGAACCCGGAATGAG 618
QY 308 AGATGAGATTAACATCACTTTAAACAATGAACTTCACTCAATCAACAATGTT 367
DB 619 AGATGAGATTAACATCACTTTAAACAATGAACTTCACTCAATCAACAATGTT 678
QY 368 TGAGTCTAATATATATTTTTCAGAGGCTTTGAAGAACATCATGATTTGCAAAATG 427
DB 679 TGAGTCTAATATATATTTTTCAGAGGCTTTGAAGAACATCATGATTTGCAAAATG 728
QY 428 GACGAAATTAATTAACCCAAATGACCAATGATTAATTCAGATCAAGTTGGTATTT 487
DB 739 GACGAAATTAATTAACCCAAATGACCAATGATTAATTCAGATCAAGTTGGTATTT 798
QY 488 GGCGTGGCTTCACTCACTTCACTTCACTGATGAAACAGATCATGCTGCACTGACG 547
DB 799 GGCGTGGCTTCACTCACTTCACTTCACTGATGAAACAGATCATGCTGCACTGACG 858
QY 548 TTAGCATTAAGTCTTTCAGAGGAGAGAGCTGTTTGAATTCAGATTCATTTATATC 607
DB 859 TTAGCATTAAGTCTTTCAGAGGAGAGAGCTGTTTGAATTCAGATTCATTTATATC 918
QY 608 AGACAGAGAACTTAATTCAGAGAGAGGTTCCAAAGAACTAATAGTTAGTTGTC 667
DB 919 AGACAGAGAACTTAATTCAGAGAGAGGTTCCAAAGAACTAATAGTTAGTTGTC 978

QY 668 TTACCAAGTATTAACAATTAAGACATTAAGAGTGAATATTTGACTGGACCAAGATCCCA 727
DB 979 TTACCAAGTATTAACAATTAAGACATTAAGAGTGAATATTTGACTGGACCAAGATCCCA 1038
QY 728 AGAGCACTTTAAGAAAGCGGCTGGAAGTCAAGCTTTAGAAATCTAAGAGAGC 787
DB 1039 AGAGCACTTTAAGAAAGCGGCTGGAAGTCAAGCTTTAGAAATCTAAGAGAGC 1098
QY 788 AATACAAACCAAGAGATCAACCACTTGAAGAGCTTGTGTGACGAGCCCAAGAA 847
DB 1099 AATACAAACCAAGAGATCAACCACTTGAAGAGCTTGTGTGACGAGCCCAAGAA 1158
QY 848 GGCGGGGCTTGGGGGAGCACTGCGAGGCTGAGCAAGCTCATTCCTGAGCTGCTATC 907
DB 1159 GGCGGGGCTTGGGGGAGCACTGCGAGGCTGAGCAAGCTCATTCCTGAGCTGCTATC 1218
QY 908 TTACAGGCTAAGTAAATGCGTATGATTTTAACTGATGAAGAACTTAGCGGTTTC 967
DB 1219 TTACAGGCTAAGTAAATGCGTATGATTTTAACTGATGAAGAACTTAGCGGTTTC 1278
QY 968 ATACAGACTAAGTCCAGAGCAAGAGCGTGAAGTGAAGAGCACTGATTAATTC 1027
DB 1279 ATACAGACTAAGTCCAGAGCAAGAGCGTGAAGTGAAGAGCACTGATTAATTC 1338
QY 1028 AATAAAGATTAATGTTCAAGAGGAGCTTGAAGCTGAGCTTGAATTCAGCT 1087
DB 1339 AATAAAGATTAATGTTCAAGAGGAGCTTGAAGCTGAGCTTGAATTCAGCT 1398
QY 1088 TACTGCTCTTCAAGAAAGATTTTGAACAGAAAGATTCACCAAGTGGAAAAATCAT 1147
DB 1399 TACTGCTCTTCAAGAAAGATTTTGAACAGAAAGATTCACCAAGTGGAAAAATCAT 1458
QY 1148 TTGATTAACAATCCCAATTTGAGATGATGTCGCAAGAAACAAGAGTGCACATTAATTA 1207
DB 1459 TTGATTAACAATCCCAATTTGAGATGATGTCGCAAGAAACAAGAGTGCACATTAATTA 1518
QY 1208 GTGTTAAGCACTGATTAAGCTGCTGATCTAATGCGAAGAAATTAATGAAGAGCA 1267
DB 1519 GTGTTAAGCACTGATTAAGCTGCTGATCTAATGCGAAGAAATTAATGAAGAGCA 1578
QY 1268 ATTCAATGATCAAAATTTAATTTAAAGTTACACAGCATGGAGTGAATGAAGAAAG 1327
DB 1579 ATTCAATGATCAAAATTTAATTTAAAGTTACACAGCATGGAGTGAATGAAGAAAG 1638
QY 1328 CAATTAATGATGAAGTGAATGACAGAACTGAAGCTTAAAGTCTTAAAGCAAAAAG 1387
DB 1639 CAATTAATGATGAAGTGAATGACAGAACTGAAGCTTAAAGTCTTAAAGCAAAAAG 1698
QY 1388 TCACAGCAGAACCCCAATGATGTTGCTGCTGTTGTAAGTAATGCGAAGCAAAATGAG 1447
DB 1699 TCACAGCAGAACCCCAATGATGTTGCTGCTGTTGTAAGTAATGCGAAGCAAAATGAG 1758
QY 1448 ATGCAATTAATAAATTAATCTGTAAGATTCAGATTCAGATTCAGATTCAGATTCAG 1507
DB 1759 ATGCAATTAATAAATTAATCTGTAAGATTCAGATTCAGATTCAGATTCAGATTCAG 1818
QY 1508 GAACCTTAAGCAACCAAGAACTGATGAGCAATTTGCTAACAAGATTTGCTTCAAGATGA 1567
DB 1819 GAACCTTAAGCAACCAAGAACTGATGAGCAATTTGCTAACAAGATTTGCTTCAAGATGA 1878
QY 1568 ACTGCAAGATGGAAGAGAGAGCTGGAAGTGAATCCCTGGAAGCTGAGATGATG 1627
DB 1879 ACTGCAAGATGGAAGAGAGAGCTGGAAGTGAATCCCTGGAAGCTGAGATGATG 1938
QY 1628 TTGCAATGATTTTAACTGATGAATGACAGCTGAGGCGAGAGTCAATCGAGAGATTTGTTG 1687
DB 1939 TTGCAATGATTTTAACTGATGAATGACAGCTGAGGCGAGAGTCAATCGAGAGATTTGTTG 1998
QY 1688 CCAAGCTCAATGAAGAGAGAGCCGCTGCTGCAAGCTGCAATTTTCAAGATGAAGAGAC 1747
DB 1999 CCAAGCTCAATGAAGAGAGAGCCGCTGCTGCAAGCTGCAATTTTCAAGATGAAGAGAC 2058

Db	1237	GCAATTACAACAGAGAGATCAACCGACCTGGAAGCACCGGTGCTGGTGAAGCAACCAAGCG	1296
Qy	846	AAGCGGGGGCCCTTGGGGGACACTGCGCCAGGGCCCTGCATGCTCATTTCTGAAGCTTGCTGA	905
Db	1297	GAGAGAGGGCCCGCGCGCACCTCGCTGGCGCCAGCATATNTCAATCCCTTAAGCTCTGCTGA	1356
Qy	906	TCTTACAGAGTCTAACTGATTAATAATGCGTAATGATTTTAAGTATGAATAACTTAGCCGT	965
Db	1357	TCTCAAGAGCGCTGACTGATGAATAATGCGAATGATTTCAATGTATATGAAGACCTGGAGCT	1416
Qy	966	TCATPACAAGACTAACTCCAGACAAAGGACGCTGAAGTGGAGCAGTCAATTGATTCAT	1025
Db	1417	GCACACGGCGGTGACCCCTGAGCAGCGCAGCGGAGGTGGCGCGCTCATCGACTACAT	1476
Qy	1026	TCATPAAAACATATATGTTCAAGGGAGCTTCAGACTGGGGGTTTGAAGTTTGAATTCAA	1085
Db	1477	CCACAAGGATGCAATGTGTCAGAGAGAGCTTCAGACTGGGGCCTGAGCTTCAGACTCAA	1536
Qy	1086	CTTACTGCTCTTCAGAGAAATTTTGCAAAACAAAAGATTCAACCAAGGTGAAAAAC	1145
Db	1537	CTTGCTGCTCTCTCTGGAAGATCTTTCATCTGAGAAGATCCACCGAGGCGGAAAGAC	1596
Qy	1146	ATTGATTTACAACTCCAAATTTGCAAGATTTGTTCCAAAGAAACAAGAGTGCACATTAAT	1205
Db	1597	GTTTATATTACAACCCACAATTTGCAAGCTGCTCCAAAGAAACAAGAGCGCGCTGAT	1656
Qy	1206	TAGTGTAAAGCCACTAGATTAATCTGGCTGTTGATCTATACGGGAAATTAATAGACAGC	1265
Db	1657	CAGCTGAAGCATTTGGATTAATGCTGCTGATCTATACCGGAGGAATTAATAGAACAGC	1716
Qy	1266	CAATTCAATTGATCAAAATCTATTTAAGTTACACACGCCATGGGCATGCAATGAGAAA	1325
Db	1717	CAACTCACTGATACGAACTCTGTTCAAGTATCTCCAGCCATGGGCATTCAGATGAATAA	1776
Qy	1326	AGCATATATGATTTGAAGTGAATGACAGAACTGAAGCCATTAATGAAGCTTACAGCAAA	1385
Db	1777	GGCATCATGATCGAGTGAATGACAGAACAAAGCTTATCTGAAGCCTTGACAGGAA	1836
Qy	1386	GCTCACAGCAGACACCCAGATAGTTGTCTGTCTGTGTTGTCAGATTAATCGAAGCAATA	1445
Db	1837	GGTGACGTCAAGCACTCAGATAGTTGTGTCTGTCTCTTGTCAAGTATCGAAGGCAATA	1896
Qy	1446	CGATGCTATTAATAAATAATCCTGTGATCAAGATTTGCCCTACCCCAAGTCAGTGTGATGCGC	1505
Db	1897	TGATCCCATCAAGAAATCTGTGTGTACAACTGCCCCACCCCAAGTCAGTGTGTGATGCGC	1956
Qy	1506	CCGAACCTTAGGCAACAGCAAACTGTCAATGCCACTTGCTCAAAAGATTGCTCAAGAT	1565
Db	1957	CCGAGCCTTGCGGCAAGAGCAAGCAAAAGTATGGCCATTTGCCACAAAGATCGCCCTGAGAT	2016
Qy	1566	GAACTGCAAGATGGAGAGAGAGCTTGGAGGTTGACATCCCCCTGAAGCTCGATGAT	1625
Db	2017	GAACTGCAAGAGGAGGAGCGAGCTCTGCGGCTGAGCAATGGCCCTGAAACCTGGCAATGAT	2076
Qy	1626	CGTTGAGCATCAATTTGTTACATGACATGACAGCTGGGCGAGAGTCAATCGCAGGATTTGT	1685
Db	2077	CGTGGGACATGACTGTATACATGACACCAAGCTGGGCGAGAGTCAATCGCAGGATTTGT	2136
Qy	1686	TGCCAGCATCAATGAAGGATATACCCGCTGTTCTTCAAGCTGTCATATTTCAAGATAGAG	1745
Db	2137	CGCCAGCATCAATGAAGGATATACCCGCTGTTCTTCCGCTGCGCTTTCAGGAACCGCG	2196
Qy	1746	ACAGAGAGCTGATAGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCGAGAGGCTTGGATAG	1805
Db	2197	GAGAGAGCTGTGATGTGTCTCAAGTGTGCTTGAAGTGTGCTTGAAGGCTTGGAGTGG	2256
Qy	1806	CTGCAATGATGATATGCTCCAGCGCTGTCAATCTGTATACCAAGACGAGTGTGGGGGAAGGCA	2316
Db	2257	CTGCAATGATGATATGCTCCAGCGCTGTCAATCTGTATACCAAGACGAGTGTGGGGGAAGGCA	2316
Qy	1866	GCTGAAAAACAATGTGAACTAAGAAAGTGCACAGTTTTCAGATTTCTTAAATTCATTTG	1925
Db	2317	GCTGAAAGCCCTGGTCAATTAATGAGTCCCAAGATTTCTTAAATTCCTCAATCAATGCTGG	2376

QY	1926	TGAGGCTTACACCCCTPAGACTAAACGGTAATTTGGTGAAGAAAAGAGTGAACCAAGATT	1985
Db	2377	GAGGAGTTTCAACCCCAAGACTGACTGTATCTGTGTGAAGAAAGCGTGTCAATGCCAGGTT	2436
QY	1986	TTTTCCTCAGTCTTGAGAAAGACTTCAGAAATCCACTTCCTGGAAACAGTTATTGATGTA	2045
Db	2437	TTTTCCTCAGTCTGGGGAAAGACTTCAGAAACCCCTTCACAGGACAGTCAATCGATGTGGA	2496
QY	2046	GCTTACCAAGACCAAAATGATATGACTTTTATCTGTAGCCAGGCTGTGAGAAATGTGTAG	2105
Db	2497	AGTCAACCAAGACCAAGTGTATGACTTTTTCATGTGTAGCCAGGCAATGAGAACGGGAG	2556
QY	2106	TGTTTCTCCACACATTCACATTCATCTATGACAACACGGGCTGAACCCACACAT	2165
Db	2557	TGTTCTCCCAACACATTCACATTCATCTATGACAAGAGTGGCTGAAGCCGACCAT	2616
QY	2166	ACACGGCTTGACCTTACCAAGCTGTCACATCTATTAACAATGCGCAGGTCATTCGTGT	2225
Db	2617	CCAGCGGTGACATCAACAGTCTGCAAGCTGGCCCTTCCTGTGGCCAGACATCCACAG	2676
QY	2226	TCCTGCTCCTTGGCAGTACGCCCAACAAGCTGGCTTTTCTGTTGGCCAGAGTATTCCAG	2285
Db	2677	CCCTGACACTTTCGACAGTATGACACAAAGCTGGCCCTTCCTGTGGCCAGACATCCACAG	2736
QY	2286	AGAGCCAAATCTGTCACTGTCAAAACCGGCTTACCTACTTA	2328
Db	2737	AGAGCCAAACTCTTCCTGTCCAAACCGGCTTACTACTCTCTAA	2779
RESULT 5			
ID	ABV89287	standard; cDNA; 2292 BP.	
XX	ABV89287;		
AC	ABV89287;		
XX	13-DEC-2002 (first entry)		
DT			
XX	Human colon cancer related cDNA SEQ ID NO 2602.		
DE			
XX	Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;		
KW	ss.		
OS	Homo sapiens.		
XX	WO200258534-A2.		
PN			
XX	01-AUG-2002.		
PD			
XX	16-NOV-2001; 2001MO-US043704.		
PF			
XX	20-NOV-2000; 2000US-0252222P.		
PR	06-FEB-2001; 2001US-0267011P.		
PR	28-MAR-2001; 2001US-0279670P.		
PR	10-JUL-2001; 2001US-0304037P.		
XX	(CORI-) CORIXA CORP.		
PA			
XX	Stolk JA, Xu J, Chenaault RA, Meagher MJ, Secrist H, King GE,		
PI	WPI; 2002-608400/65.		
XX	P-PSDB; ABP67995.		
DR			
XX	New isolated tumor colon polynucleotide and polypeptide, useful for the		
PT	diagnosis, prevention and/or treatment of cancer, in particular colon		
PT	cancer.		
XX	Claim 1, SEQ ID NO 2602; 266pp + Sequence Listing; English.		
PS			
XX	The invention relates to a human colon tumour expressed polynucleotide		
CC	(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of		
CC	2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)		
CC	complements of (i); (iii) at least 20 contiguous residues of (i); (iv)		

Db	1703	GACTACCCCAAAATCAATAAAGTACAGAAATCCAGCTGCATTGTTGAGCTATACG	1762
QY	1381	CAAAAGGTCAACAGCAGACACCCAGATAGTGTCTGTCTTGTTCAGATATGGAAGAC	1440
Db	1763	CAATATGTTGATCTCTGATGTTTCAGCTGTATGTGATTCCTGCTTCTATCAAGAGACC	1822
QY	1441	AAATACGATGCTATTAAAAATACCTGTGACGATGTGCCCTACCCCAAGTCAGTGTG	1500
Db	1823	TATTATATTCCTATTAAAAATTTTGTAGCTCAGACTGCCCAATCCCAAGCCAAATGTGTG	1882
QY	1501	GTGGCCCCGAACCTTAGCCAAACAGCAACCTGTATGCGCATTCCTACAAAGATTGCCCTA	1560
Db	1883	CTTGCTCGGACCTTGAATTAACAGGGCATGATGATGTCCCAACAAAGTCGCTATG	1942
QY	1561	CAGATGAATCTGCAAGATGGGAGGAGAGCTCTGAGAGGGGTGACATCCCCCGAAGCTGTG	1620
Db	1943	CAGATGACTTGGCAAGCTCGGAGGGAGAGCTGTGGGCTGTGAAAAATCCTTTAAAGTCCCTG	2002
QY	1621	ATGATCGTTGGCATCGATTGTTATTCATGACATGACAGACTGGGCGGAGTCAATGCAAGA	1680
Db	2003	ATGTTGTTCCGTAATGATGATGCTGTAAAGATGCACTCAGCAAGACGTAATGTTGTGGA	2062
QY	1681	TTTGTTCGACATCAATGAAGGATGACCCCGTGTCTCAGGCTGCATATTCAGAGAT	1740
Db	2063	TGCGTGGCAGATGTGTAACCCCAAGATCACCAAGGTGTTTTCCGCTTATCCTTCAGAGA	2122
QY	1741	AGAGGACAGAGAGCTGTATGATGGGCTCAAAAGTCTGCTCGCAAGCGGCTCTGAGGGCTTG	1800
Db	2123	ACAAATGACTGATGTTGACAGATTGCTTGAAGTTTCAATGACTGGAGCACTCAACAAATGG	2182
QY	1801	AATAGCTCAATAGATACATATCCCAAGCCGATCATCTGTATACCCGATAGCGATAGAGAC	1860
Db	2183	TACAAGTCAATATCATGATTGTCCAGCAAGATAATGTGTACCTGTGTGTGTAGGGGAT	2242
QY	1861	GGCCAGCTGAAAAACCTGTGAATCAACAAATGCGACAGTTTGTGATTTGATCTTAAATCC	1920
Db	2243	GGTACGCTGGAAACCTTATTTGATATGATATCCCAAGCTGCTGAGCAGGTGTGGCAGAA	2302
QY	1921	ATTGTGTAAGGTTTCAACCTTACACTTAACGTTAATTTGTGTGAAAGAAAGTGAACAC	1980
Db	2303	TCCGCGTCAATATACAGCTCAAGACTGTGCGTGTATGTGTGTCAGAGAAAGATGATGCCA	2362
QY	1961	AGATTTTGTTCCTCAGTCTGAGAGGAAAGCTTGAGATCACTTCCGGAACAGTTATTGAT	2040
Db	2363	CGATTCTTTTACCGAAATGAAACCGCACTGTACAGAACCCCACTTGGCAGCTGTGTGAT	2422
QY	2041	GTAGAGGTTTACAGACACGAATGTATGACTTTTATCGTAGCCAGGCTGTGAAGT	2100
Db	2423	TACGAAGCAACAGTAAAGAAATGATATGACTTTTATGATACAGCAAGTGTGCTGCGG	2482
QY	2101	GGTATGTTTTCCTCCACATTTACATATGTCATCTATGCAACAGGCGGCTGAAGCAGAC	2160
Db	2483	GGAAAGTTTATGTCACCTACATATATGTCATCTATGATGACAGAGCTTGAAGCCGAC	2542
QY	2161	CACATACAGCGCTTGACTTACAAAGCTGTGCAATCTATTAACAATGCGCCAGGTCAAT	2220
Db	2543	CATATCAGAGACTTACATTAATATTTGTGCAACCTGTACTACAACTGGCGCGGCAATATC	2602
QY	2221	CGTGTTCGTCTCTTCCAGTACGCGCCCAAGCTGTGCTTTCTGTGTGGCAGAGTAT	2280
Db	2603	AGGTGCCAGACCAATGATCAGTATGTCTCAACAGCTCACTTCTGTGTGCACAAAAGCAAT	2662
QY	2281	CAAGAGAGCAATCTGTCACTGTCAAAACCGGCTTACTACTCTCTA	2348
Db	2663	CATTAAGAACCACTGTGGAATTAGCAACATCTTCTTACCTGTGA	2710

DT	26-FEB-2003	(first entry)
XX		
DB	Human mRNA sequence encoding sequence.	
XX	Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;	
XX	gene therapy; nutritional supplement; wound; burn; ulcer;	
KW	Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;	
XX	autoimmune disorder; inflammation; vulnerability; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
FN	WO200231111-A2.	
PD	18-APR-2002.	
XX		
PF	11-OCT-2001; 2001WO-US027760.	
PR	12-OCT-2000; 2000US-00687527.	
XX		
PA	(HXSE-) HXSEQ INC.	
PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;	
XX	Xue AJ, Yang Y, Wehrman T, Drmanac RT;	
XX	WPI; 2002-426278/45.	
DR	N-PSDB; ABP43865.	
XX		
PT	New polypeptides and their encoded proteins, useful as nutritional	
PT	sources or supplements, or in gene therapy, particularly for treating	
PT	wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or	
PT	inflammation.	
PS		
XX	Claim 1; SEQ ID # 322; 357bp + Sequence Listing; English.	
XX		
CC	The invention relates to 446 newly isolated polynucleotide sequences. The	
CC	activity of polynucleotides of the invention may be described as,	
CC	vulnerary, neuroprotective, immunomodulator, cytostatic and anti-	
CC	inflammatory. Compositions comprising nucleic acids of the invention are	
CC	useful for treating a mammalian subject, or as nutritional sources or	
CC	supplements. These are useful in gene therapy, particularly for treating	
CC	wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,	
CC	amyotrophic lateral sclerosis, autoimmune disorders, cancer or	
CC	inflammation. The nucleic acids and polypeptides are also useful in	
CC	diagnostic and research methods. The sequences given in records ABQ6788-	
CC	ABQ67233 represent polynucleotides of the invention. NOTE: The sequence	
CC	data for this patent did not form part of the printed specification, but	
CC	was obtained in electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
XX		
XX	Sequence 2009 BP; 595 A; 411 C; 458 G; 545 T; 0 U; 0 Other;	
QY		
Query Match	19.7%; Score 458.4; DB 6; Length 2009;	
Best Local Similarity	57.2%; Pred. No. 5.3e-126;	
Matches 844; Conservative	0; Mismatches 661; Indels	9; Gaps
QY		3;
DB		
765	CTTCTTAAGATCTACAGAAAGCAATACACCAAGAGATCAACGACTTGAAGCGCTGT	824
15	CTAGGTGGATTCTACACAGACGATAGATTAATCTGTATCGACTGAATCGCCCAT	74
QY		
825	CTTGTGACGCGCCCAAGAGAAGCGCGGGCCCTGGGGGACACTGCCAGGCGCTGCCAT	884
DB		
75	GCTTGTTAGTCTGTAAAG--AAGAAAGAAAGCAACAGTAGAGCCTCAGTCGCGCA	131
QY		
885	GCTCATCTCTGAGCTGTGTAATCTTACACGCTCTACATGATTAATGCGTATGATTTAA	944
DB		
132	CCTGATACCTGAGCTCTGCTTTCTTACAGGGGTGACTACAGGACACATCTGATTTCCA	191
QY		
945	CGATGATGAAGACTTAGCGCGTTCATACACAGACTTAATCTCAAGACAAAGCGCTGAAGT	1004
DB		
192	GCTGATGAAGGCTGTGGCTGAAGAAAGCAGTCTCAGTCTTCAGGCGCGACAGCGCT	251
QY		
1005	GGAAGCACTCATTTGATTAATTCATTAAGATATGTTCAAGAGGAGCTTCGAGACTG	1064

CC human partial Hiwi-Hy cDNA appearing as AC62541; (b) having the stem
 CC cell maintenance factor protein coding nucleotide sequence of a
 CC polynucleotide of (a); or (c) having the mature stem cell maintenance
 CC factor protein coding nucleotide sequence of (a). Also include are an
 CC isolated polynucleotide encoding a polypeptide with stem cell maintenance
 CC factor activity that hybridises under stringent conditions to the
 CC complement of (or has a sequence identity greater than 90% to AC62541),
 CC an expression vector comprising the Hiwi-Hy cDNA, a host cell genetically
 CC engineered to contain or express the Hiwi-Hy cDNA in operative
 CC association with a regulatory sequence that controls expression of the
 CC polynucleotide in the host cell, the partial Hiwi-Hy stem cell
 CC maintenance factor appearing as AB061932, an anti-Hiwi-Hy antibody,
 CC detecting Hiwi-Hy in a sample (comprising contacting the sample with a
 CC compound that binds to and form a complex with the polypeptide),
 CC identifying the binding compound and modulating cell survival and/or
 CC death (comprising contacting the cell with Hiwi-Hy for a time sufficient
 CC to modulate the activity). The methods and compositions of the present
 CC invention are useful for the diagnosis and treatment of disorders such as
 CC diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease
 CC and organ transplantation. They can also be used for re-engineering
 CC damaged or diseased tissues, manufacture of pharmaceuticals and in the
 CC development of biological-based sensors. The present sequence is the full
 CC length cDNA encoding the human Hiwi-Hy protein

XX Sequence 2009 BP; 595 A; 411 C; 458 G; 545 T; 0 U; 0 Other;

Query Match 19.7%; Score 458.4; DB 8; Length 2009;

Best Local Similarity 57.2%; Pred. No. 5,3e-126; Mismatches 661; Indels 9; Gaps 3;

DB 765 CTCTTGAATTAATACAGAGCAATACCAAGAGATCAAGAGCTTGAAGAGCTGT 824
 15 CTAGTGGATTAATACAGAGCAATACCAAGAGATCAAGAGCTTGAAGAGCTGT 74
 QY 825 CTGTGTACAGAGCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 884
 75 GCTGTGTAGTGTGTAAAG--AAGAAAGAAATGACACAGAGAGAGAGAGAGAG 131
 DB 885 GCTCATCTGAGCTGTGTATCTTACAGAGCTTGAAGAGAGAGAGAGAGAGAG 944
 132 CCGATACCTGAGCTGTGTATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 191
 QY 945 CGTGTGAAAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
 192 GCTGTGAG 251
 DB 1005 GAG 1064
 252 GAG 311
 QY 1065 GAG 1124
 312 GAG 368
 DB 1125 GATTAACAG 1184
 369 AATTAATTAATGAG 425
 QY 1185 AACAG 1244
 426 TATTCGAG 485
 DB 1245 GAG 1304
 486 GAG 545
 QY 1305 CATGGAG 1364
 546 CATGGAG 605
 DB 1365 CTTAAG 1424
 606 TGTTCGAG 665

QY 1425 AAGTATCGAAG 1484
 DB 666 TTCTATACAG 725
 QY 1485 CCGAG 1544
 DB 726 CCGAG 785
 QY 1545 TACAAG 1604
 DB 786 CACAG 845
 QY 1605 CCGAG 1664
 DB 846 ACCTTAAAG 905
 QY 1665 GAG 1724
 DB 906 CAG 965
 QY 1725 CTGATATTCAG 1784
 DB 966 CTGATATTCAG 1025
 QY 1785 GAG 1844
 DB 1026 AGCAG 1085
 QY 1845 CGATGAG 1904
 DB 1086 TGTGTGTGAG 1145
 QY 1905 GATGTGTGAG 1964
 DB 1146 GAG 1205
 QY 1965 GAAAG 2024
 DB 1206 GAG 1265
 QY 2025 TGAAG 2084
 DB 1266 TGAAG 1325
 QY 2085 CAG 2144
 DB 1326 CAG 1385
 QY 2145 CAG 2204
 DB 1386 CAG 1445
 QY 2205 CAG 2264
 DB 1446 CAG 1505
 QY 2265 TGTTCGAG 2324
 DB 1506 GAG 1565
 QY 2325 CTAA 2328
 DB 1566 GTGA 1569

RESULT 9

ADA53334
 ID ADA53334 standard; cDNA, 3620 BP.

ADA53334;

AC AC
 XX XX
 DT 20-NOV-2003 (first entry)

XX Human coding sequence, SEQ ID 902.
 XX Cyostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 XX Gene Therapy; human; secretory protein; membrane proteins; cancer;
 XX inflammatory disease; osteoporosis; neurological disease; gene; ss.
 OS Homo sapiens.
 XX EPI293569-A2.
 XX 19-MAR-2003.
 XX 21-MAR-2002; 2002EP-00006586.
 XX 14-SEP-2001; 2001JP-00328381.
 XX 24-JAN-2002; 2002US-0350435P.
 XX (HELI-) HELIX RES INST.
 XX (REMS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Sugiyama T, Otsuka T, Wakamatsu A, Sato H, Ishii S,
 XX Yamamoto Y, Isono Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 XX Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;
 XX MPI: 2003-395539/38.
 XX P-PDB: ADA54973.
 XX New polynucleotides encoding full-length polypeptides, e.g. secretory
 XX and/or membrane proteins, useful for developing medicines for diseases in
 XX which the gene is involved, or as target molecules for gene therapy.
 XX Claim 1; SEQ ID NO 902; 205bp; English.
 XX The present invention relates to novel human secretory or membrane
 XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 XX ADA54071). The coding sequences are useful in the gene therapy of
 XX diseases caused by abnormalities of the proteins, e.g. cancer,
 XX inflammatory diseases, osteoporosis or neurological disease.
 XX Sequence 3620 BP; 947 A; 847 C; 946 G; 880 T; 0 U; 0 Other;
 Query Match 18.7%; Score 435.2; DB 7; Length 3620;
 Best Local Similarity 52.0%; Pred. No. 6,9e-119;
 Matches 1153; Conservative 0; Mismatches 1043; Indels 22; Gaps 7;

QY 544 GACGTTAGCCATAAAGCTCTTGGAAGTGAAGTCTGTTTGAATTCAGTTCACTTTAT 603
 DB 1266 GATGCTCCCAATAGGTCATTCGGAATACCTGTGTCTGGAATGTCATGATGATGATTTAT 1325
 QY 604 CATTGACAGAAAGACATTAATTTCAAGAACAGATTTCAGAACATTAAGTTAGTT 663
 DB 1326 CAGCAGATTAAGAAC--ACTTCAGAGTAGTGTACTTAAGCTCTGTTGGCAATAT 1382
 QY 664 GTTCTTACCAAGTATTAACATTAAGACATACAGATGATGATATTTGACTGGGACAGAT 723
 DB 1383 GTTATCACCAGATATTAACATGATGATGATGATGATGATGATGATGATGATGAT 1442
 QY 724 CCAAGAGACCTTTTAAGAAAGCCGCTGGAAGTCAAGCTTGTAGATTAATTAATCAAG 783
 DB 1443 CCAAGAGATAGCTTCAAGATGCTGATGATGGAAGAAAGATCAATTTGGAATTAATCAAG 1502
 QY 784 AAGCAATTAACCAAGATTAACCAAGATTAACCAAGATTAACCAAGATTAACCAAG 841
 DB 1503 AAAAATTAAGGATTAACCAAGATTAACCAAGATTAACCAAGATTAACCAAGAT 1562
 QY 842 -AGAGAGGCGGCGCTGCGGGGACATGCGAGCGCCATGCTCATTCCTGAGCTC 900
 DB 1563 GAGAGACAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1622
 QY 901 TGCTATCTTACAGGCTTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 960
 DB 1623 TCTTTATAGCGGAATCCAGAGAAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1682
 QY 961 GCCGTTATTAACAGTAACTCCAGAGCAAGGACGAGAGTGGAGAGTCACTTATGAT 1020
 DB 1683 GCTCAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1742
 QY 1021 TACATTAATTAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 1080
 DB 1743 AGAATTAAGAAAG 1802
 QY 1081 TCCAACTTACTGCTCTTCAAGAGAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1140
 DB 1803 AAGATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1859
 QY 1141 AAAAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1200
 DB 1860 AATACCTGTTTATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1919
 QY 1201 TTAATTAAGTAAAG 1260
 DB 1920 TCCATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1979
 QY 1261 GAGGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1319
 DB 1980 CAGGCTGAG 2039
 QY 1320 GAGAAAGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1376
 DB 2040 GAGGCCACCGCGGCTGGAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 2099
 QY 1377 ACAGCAAAAGTCAAGAG-----ACACCAGATGATGATGATGATGATGATGATGAT 1430
 DB 2100 TCAATCAAGTAAAG 2159
 QY 1431 TCGAAGAGCAAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1490
 DB 2160 ACGTATTAATTAAG 2219
 QY 1491 TCAAGTGTGAG 1550
 DB 2220 CAGAGTGTGAG 2279
 QY 1551 GATTCGCTTACAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1610
 DB 2280 GATTTTACTTCAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2339

QY 1611 GAAGCTGTGATGATCGTTGGCATCGATTGTTACATGATGACAGCTGGGCGAGATC 1670
 DB 2340 GAAACAGTTAATGTTGATCGGAGATGATGTTTACATGACCCCGATGAGCATGCGCTC 2399
 QY 1671 AATGCGAGATTTTGTGCGACATCATGAAAGGATGACCGCTGGTTCTCAGCTGCAT 1730
 DB 2400 CGTGGTGGCTTCGCGACATCATCTCACTCAAAATGATATCCCGGATGTG 2459
 QY 1731 ATTTCAGATGAGAGACAGAGAGCTGTGATGAGGCTGCAAACTGCGCTGCAAGGCGCT 1790
 DB 2460 GTTCCAGATGCGGATCGAGAGATGTTGACAGCTGAGCTGATGCTGCGGCTCTT 2519
 QY 1791 GAGGCGCTTGAATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATG 1850
 DB 2520 AAAAAAGTTTATGAGTGAACCACTGCTACAGAGAAAGATGATGATGATGATGATGATG 2579
 QY 1851 CGTAGAGAGCGGCGCTGAAACCTGCTGATGATGATGATGATGATGATGATGATGATG 1910
 DB 2580 AGTGTCTATGCGCACTGAAAGAGCTGCAACTGATGATGATGATGATGATGATGATGATG 2639
 QY 1911 TCTAAATCCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1970
 DB 2640 TTTTAAAGCTTTTGAGAA---TTTACAGCCCAAGATGATGATGATGATGATGATGATGATG 2696
 QY 1971 AGTGAACACCGATTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2030
 DB 2697 AATCAGTACTATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2756
 QY 2031 AGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2090
 DB 2757 TGTGTATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2816
 QY 2091 TGTGGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2150
 DB 2817 TGTACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2876
 QY 2151 GAAAGCAGACCAACATGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2210
 DB 2877 GAGCCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2936
 QY 2211 AGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2270
 DB 2937 TGGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2996
 QY 2271 CGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2328
 DB 2997 ACACATCTTGCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3054

RESULT 10
 AAF21880
 ID AAF21880 standard; DNA; 423 BP.
 AC AAF21880;
 XX
 DT 27-MAR-2001 (first entry)
 DE
 XX Human breast and ovarian cancer associated antigen gene SEQ ID 267.
 XX
 KM Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KM neurotropic; neuroprotective; antiviral; antiallergic; hepatocarcinoma;
 KM antidiabetic; antineoplastic; antitumor; antitumor; antitumor;
 KM antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KM Addison's disease; allergy; autoimmune haemolytic anaemia;
 KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KM cardiovascular disorder; wound healing; neurological disease; ds.
 OS Homo sapiens.
 XX
 XX WO200055173-A1.
 PN
 XX
 PD 21-SEP-2000.

XX
 PF 08-MAR-2000; 2000MO-US005881.
 XX
 FR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR P-PSDB; AAB58977.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention, treatment
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
 PT neurological diseases.
 XX
 PS Claim 1; Page 686; 1299pp; English.
 CC
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic;
 CC neuroprotective; antiviral; antiallergic; hepatocarcinoma; antidiabetic;
 CC antineoplastic; antitumor; antitumor; antitumor; antitumor;
 CC antifungal; antiparasitic and cardiac activity. The polynucleotide and
 CC protein sequences are used in the diagnosis of cancer, particularly
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
 CC and antagonists may also be used in the diagnosis, prevention and treatment
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases
 XX
 SQ Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 0 U; 4 Other;
 Query Match 17.0%; Score 395.4; DB 3; Length 423;
 Best Local Similarity 98.8%; Pred. No. 1.6e-107;
 Matches 404; Conservative 4; Mismatches 0; Indels 1; Gaps 1;
 QY 1474 GATTGCGCTACCCCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1533
 DB 14 GATTGCGCTACCCCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 73
 QY 1534 ATGGCCATTGCTACAAAGTGTGCTTCAAGATGATGATGATGATGATGATGATGATGATG 1593
 DB 74 ATGGCCATTGCTACAAAGTGTGCTTCAAGATGATGATGATGATGATGATGATGATGATG 163
 QY 1594 AGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1653
 DB 134 AGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1693
 QY 1654 ACAGCTGGGAGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1713
 DB 194 ACAGCTGGGAGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 253
 QY 1714 TGGTTCACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1773
 DB 254 TGGTTCACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313
 QY 1774 TGGCTCAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1833
 DB 314 TGGCTCAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373
 QY 1834 ATGCTGATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1892
 DB 374 ATGCTGATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421

QY	1678	GGATTTGTTGCCACATCAATGAAGGATATACCCGCTGGTCTTCACAGCTGCATATTTCCAG	1737
Db	1195	GGGCTTCGTGGCCAGCATAAATCTCAACATCACCAAATGGTACTGAGGGTGGTCTCCAG	1254
QY	1738	GATAGAGGACAGAGNGCTGTAGATNGGGCTCAAACTCTGCCTGCAGACGAGCTGTAGGGCT	1797
Db	1255	ATGCCACATCAGGAGATTTGTGACAGAGCTGAAGCTCGCTGGTGGGTTCTCTTGAAAAAG	1314
QY	1798	TGGAATAGCTGCATATGATACATGCCACGCGGATCATCGGTATACCGCATGGCGTAPAGA	1857
Db	1315	TATTTATAGGTGAACCATTTGTCTCTCCAGAGAAATTTGTGGTGTATCCAGATATGAGTGTCT	1374
QY	1858	GACGCGCAGCTGAAAACACTGTGTAGACTACGAACTGCACAGTTTGGATTGTCTPAAA	1917
Db	1375	GATGCGCAGCTAAAAGACAGTTGCCACATCAACAGATCCCTCAGCTGCAGAAAGTGTTTGAA	1434
QY	1918	TCCATTGGTAGAGTTCAACCTTAGACTAAGCTAATTGTGTGGAAGAAAAGAGTGAAC	1977
Db	1435	GCCTTTGATA--ACTACACCCCAAGATGGTGGTGTGTAGTTGATTCAGAAATAATCAGC	1491
QY	1978	ACCAATTTTTTGTCTCAAGCTGTGAGGAAAGATTCAGAAATCCACTTCTGTGAACAAGTATT	2037
Db	1492	ACCAATCTGTACCTTGCTGTCTCTGTATCACTTGTCAACCCCTCCCGGAGCTGTGGTT	1551
QY	2038	GATGAGAGGTTTACAGACAGAAATGTATGACTTTTTTATCGTAGCCAGGCTGTAGA	2097
Db	1552	GATCATACATTAACCAAGCTGTGAGTGGGTGATTTCTACCTTCTTGCCATCATGTGCGA	1611
QY	2098	AGTGTATGTATTTCTCCACACATTAACATATCATATATGACCAACAGCGGCTGAAGCA	2157
Db	1612	CAGGGCTTGGCATACTCAACACATCACTACTGTGTCTTGAAACATGCAAAATCTBAGCCT	1671
QY	2158	GACCACATACAGCGCTTGACCTAACCTACAGCTGTGCACATCTATATACATGCGCAGGTCT	2217
Db	1672	GATCACATGACAGAGGTGATCTTCAAACTATGCCACATGATCTGAAATTTGGCCTGTACC	1731
QY	2218	ATTGCGTTCCTGGCTCCTTGGCCAGTAGGCCCAAGAGTGGGCTTTCTTGTTGGCCAGGT	2277
Db	1732	ATCCGATTTCCAGCTCTCTTGCAAGTATGCCCCAGAGCTAGCTTCTGTGCCGACAGATT	1791
QY	2278	ATTCAACAGAGCCAAATCTGTCACTGTCAAAACCGCTTTACTACTCTTAA	2328
Db	1792	TTGACATCATGACCCAGCATTCAGCTGTGTGGAACTGTGTTCTCTGTAA	1842

RESULT 12	
AB219511	
ID	AB219511 standard; cDNA, 367 BP.
XX	
XX	
AC	AB219511;
XX	
DT	23-JAN-2003 (first entry)
XX	
DE	Group III cDNA cancer related clone SEQ ID NO:1937.
XX	
XX	Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
KW	immune response; virology; immunology; microbiology; molecular biology;
KW	recombinant DNA technology; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200278516-A2.
XX	
PD	10-OCT-2002.
XX	
PF	28-MAR-2002; 2002MO-US010421.
XX	
PR	30-MAR-2001; 2001US-0280255P.
PR	28-AUG-2001; 2001US-031563P.
PR	09-JAN-2002; 2002US-0347313P.
XX	
XX	
PA	(CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI: 2003-058387/05.
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and in
XX vitro, immunology, microbiology, molecular biology and recombinant DNA
XX technologies.
XX
XX Claim 1; SEQ ID NO 1937; 207pp; English.
XX
XX ABO17575 to ABO20506 represent isolated polynucleotide (I) sequences, and
XX CC ABO54446 to ABO54472 represent protein (II) sequences, from the present
XX CC invention. (I) and (II) have cytostatic activity and can be used in gene
XX CC therapy and vaccines. (I), (II), antibodies and compositions from the
XX CC present invention are useful for diagnosing, preventing and treating
XX CC cancer, which expresses CT or CP mRNA antigens. They are useful for
XX CC stimulating immune response. They can also be useful in vitro, in
XX CC immunology, microbiology, molecular biology and recombinant DNA
XX CC techniques. N.B. The sequence data for this patent did not form part of
XX CC the printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 367 BP; 118 A; 77 C; 91 G; 81 T; 0 U; 0 Other;

Query Match	15.7%	Score 365.4	DB 7	Length 367
Best Local Similarity	99.7%	Pred. No. 1,4e-96		
Matches 366	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1231	AAAGTTACACAGCGCATGGGCATGCAATGTGAGAAAAGCAATTAATGATTGAATGATGAC	1350	
DB	1	AAAGTTACACAGCGCATGGGCATGCAATGTGAGAAAAGCAATTAATGATTGAATGATGAC	60	
QY	1351	AGAACTGAAGCCCTACTTAAGTCTTACAGCAAAAGTCAACGAGACACCCGATGTT	1410	
DB	61	AGAACTGAAGCCCTACTTAAGTCTTACAGCAAAAGTCAACGAGACACCCGATGTT	120	
QY	1411	GCCTCTGTTGTCAAGTAATCGAAGAGCAAAATACGATGCTATTAAAAATACCTGTGT	1470	
DB	121	GCCTCTGTTGTCAAGTAATCGAAGAGCAAAATACGATGCTATTAAAAATACCTGTGT	180	
QY	1471	ACAGATTGCCCTACCCCAAGTCAGTGTGTGTGGCCCGAACTTAAGCAAAACACAAACT	1530	
DB	181	ACAGATTGCCCTACCCCAAGTCAGTGTGTGTGGCCCGAACTTAAGCAAAACACAAACT	240	
QY	1531	GTCAATGAGCCATTGTACAAAGTTGCCCTACAGATGAACCTGCAAGATGGAGAGAGCTC	1590	
DB	241	GTCAATGAGCCATTGTACAAAGTTGCCCTACAGATGAACCTGCAAGATGGAGAGAGCTC	300	
QY	1591	TGAGAGGTGACATATCCCTTAAGTCTGATGATCGTTGACATCGATGTTTACATGAC	1650	
DB	301	TGAGAGGTGACATATCCCTTAAGTCTGATGATCGTTGACATCGATGTTTACATGAC	360	
QY	1651	ATGACAG 1657		
DB	361	ATGACAG 367		
RESULT 13				
ABZ2005				
ID	ABZ20005	standard; cDNA; 367 BP.		
XX	ABZ20005;			
XX	23-JUN-2003	(first entry)		
DT				
XX	Group III cDNA cancer related clone SEQ ID NO:2431.			
DB				
XX	Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;			
XX	Immune response; virology; immunology; microbiology; molecular biology;			
KW	recombinant DNA technology; gene; ss.			
XX				

[illegible]

Db	1138	TGAGCAGCCTGAAAGCTATGCCTGTGGGCTTCCTTAAAAAGCTTTATGAGGTGAACACT	1197
Qy	1817	ACATGCCAGGCGGANTCATCGTGTATCCGCGGATGGCGGTAGAGACGCGCAGCTGAAAAACAC	1876
Db	1198	GCTACACAGAAATATGTGGTGTACCGTGATGAGGTGTCTGATGGCCAACTGGAAGACAG	125
Qy	1877	TGTGAATACGAAGTGCACAGTTTGTGATTTCTCAAAATTCATTTGGTGAAGTTTACA	1936
Db	1258	TTGGCAACTATGAAGATTCCTCAACTACAGAAAGTGTTTGAAGCTTTTGAGA--ATTATC	1314
Qy	1937	AACCTAGACTAACGGTAAATGTGGTGAAGAAAGATGAACACCAAGATTTTGTCTAGT	1996
Db	1315	AGCCCAAGATGTGGTGTGTGTGAATTGAGTTGAGAAAGAAATCAGTACTATATCTTGAGCTG	1374
Qy	1997	CTGAGAGAAGACTTCAGAAATCCACTCTCTGGAAACGATTAATGATGATAGGTTACCAAG	2056
Db	1375	CTCCTCAGAACTTTGTATCTCCACTCTCTGGAACTGTGTGATATATACATTAACAAGT	1433
Qy	2057	CAGAAATGATGACTTTTATATCTGAGCCAGGCTGTGAGAAAGTATAGTGTCTCCCA	2111
Db	1435	GTAAGTGGGTGAATTTCTATCTTCTTGCCCATCATGTACGCGAGGCTGTGCATTTCTTA	1491
Qy	2117	CACATTACAAATGTCAATCTATATGACAACAGGCGCTTAACCCAGACCAATACAGCGCTTA	2177
Db	1495	GCGATTATATCTGTGTCTTCTCAACCCGAAACCTGAGCCCTGATATATGCAAGAGCTGA	1551
Qy	2177	CCTAACAAGCTGCGCAATCTATATCAACTGGCGAGGTGTCAATGTGTCTGTCTCT	2231
Db	1555	CTTTCAAACTGTGCCAATGTACTGGAATTTGGCTGTGGACCAATCAGAGTTCCAGCTCTT	1611
Qy	2237	GCCAGTACGCCCAACAAGCTGAGCTTTCTTGTGTTGCCAGAGTATTCACAGAGGCCAATC	2291
Db	1615	GGAAGTATGCCCAAGAGCTAGCTTTCTCTCAGACACATCTTGATCATGAGCCAGCCA	1671
Qy	2297	TGTCACTGTCAAAACCGCCTTTACTACCTCTAA	2328
Db	1675	TCCAGCTGTGCGAGAACTGTGTCTTCTCTGTGA	1706
RESULT 15			
ABL10571			
ID	ABL10571	standard; cDNA; 2838 BP.	
XX	ABL10571;		
AC			
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.		
XX	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
XX	WO200171042-A2.		
PN			
FD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
XX			
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li FWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
XX			
DR	P-Psdb; ABB66468.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signaling and cell-cell		
PT	interactions.		

XX Claim 1, SEQ ID NO 26195; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB101840-AB16175) and the encoded proteins (AB571737-AB572072). The sequence data for this patent does not form part of the CC printed specification, but was obtained in electronic format directly from WIPRO at http://wipro.int/pub/publ/published_pcc_sequences

XX Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 U; 0 Other;

Query Match 13.3%; Score 310.6; DB 4; Length 2838;

Best Local Similarity 49.1%; Pred. No. 1.4e-81; Mismatches 1134; Conservative 0; Mismatches 1129; Indels 46; Gaps 10;

52 AAAAAGGTTCTTCAGCATTTATAGATTAAAGCTTAACCACTTCCGCTGACATCC 111
431 AAGAAAGAGTGGTGGCACTCATATTACCGTAGAGCAAACTATTTTAAAGTATTAAG 490
112 CGTCCCAAGTGGGCTTATATAGATTCATGACTATTAACCACTGATGGAAGCGAGA 171
491 CGTCAAACTGAGCACTACACAGTACCGGTGATTTTACGCTGATGAGGCTGACA 550
172 AGACTCCGTTCAAGCTCTCTTTTCAACAAGAGATCTAATGAAAGTGCATGCTTT 231
551 CGACTCGACGGCTTTCTTGATGAACTAAAGGATCTCGGGC--GGCTCATCTTGG 608
232 GATGAGCACTATTTATTTTACCTAAAGACTACAGCAA-----AGTTA 277
609 ACGGAACCAATATGTTTGCATCAATCAAGTTCAAACTGTTCAAGATAGCCCTATGTT 668
278 CTGAATTTTATAGTAAAGCCGGAATGAGAGATGAGATTAAGATCACTTTTAAACA 337
669 TGGAATCTGTTACGAAGAGTGGTGGGAAACATTAATTAAGATCAAGGCTGTTG 728
338 ATGAATCTTCACTTACATCAACAATGTTTGGATGTTTATTAATTTTACAGAGGC 397
729 GATCTGTGCAATCTACAGATGCGAGCAATTTCAAGTCTTATATCTCAATACGCAAGG 788
398 TTTTGAATTCATGATTTTGAACAACAATTTGACGAATATTAATTAACCAATAGACCA 457
789 CCAATGAGGCTTACATTAAGTGAAGTGTCTCGGCTCTACTCAATCTCAAGCTAAGA 848
458 TTGATATTCAGATGCAAGGTTGATGTTGGCTGGCTTCACTACTTCACTCTTCACT 517
849 TTAATTTGAGAAATTTCCCATGCAATTAATGCTGGCTATCAAGCTTGAATCCCAAGC 908
518 ATGAATAACAGCATCAAGTCTGCACTGAGATTAAGCTTAAGCTTGAATGAGAGCTG 577
909 ACGAAATATGATTAATTAATTTTCTCCGATATGCGCAAGTTATGCAACTGAGACT 968
578 TTTTGAATTCATGTTCAACTTTTATCATCAGACAGAAAGCAATTAATTTCAAGAACAG 637
969 TGATACATATTTTGTCCGATGCTATTGCGAGACGATGACATTAATCAATTAAGC 1028
638 TTTTCAAAAGACATTAAGTGTGTTGTTCTTACCAAGTAAACATTAAGATACAGAG 697
1029 GT-----GCAATTAAGGATGATGATTAATTAATTAATTAATTAATTAATTAAT 1082
698 TGATATATTTGATGCTGAGACAGAAATCCCAAGAGCACTTTAAGAAAGCCGCTGCTG 757
1083 TTGACATGTCGATCTTCAATGAGCGCATTTGCAAAATTT--AAAACCAATGACGGTG 1139
758 AATGTCAGCTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 817
1140 AGATTTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1199
818 AGCTGTCTTGTGACAGCCCAAGAGAGGCGGGCTCTGGGGGAGACATGCGCAGGCG 877

DB 1200 AACCTGTGATCATGTCTC--GTCCAGACAGCAAAATATTTGGTGGAAATGACAGGCC 1258
878 CTGCGATGCTCATCTCTAGAGCTCTGATATCTTACAGGCTTAATGATTAATGATGCTAATG 937
1259 ATA--ATGATCATTTCCGAGCTGACAGGCTTACGGAATGAAGAGACGATATGCGCTG 1316
938 ATTTTAAGTATGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 997
1317 ACTTTAGAGCTTGAAGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1376
998 GTGAAGTGGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1057
1377 AAGCCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436
1058 GAGACTGGGGTTGAGCTTGAATTCACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1117
1437 AGTCTGGAACATGAGAGCTGAGCTCGCTTGGTGAAGATTCAGACAGCGGTGTTGCCAC 1496
1118 CAGAAAGATTCACCAAGCTGAGAAACATTTGATTAATTCATCAATTCAGATTTGAGATGCT 1177
1497 CGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1556
1178 CCAAGAAACAGAGTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1237
1557 CCAATGAGTTTGGACCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1616
1238 TCTATACGCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1297
1617 TCACACCGAGTCCAAATTCGCGGAGACCGAGATTTGCGAGATGATGATGATGATGATGATGATG 1676
1298 CACAGCATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1354
1677 CCAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1736
1355 CTGAAGCTTAAAGTCTTACAGCAAAAGTCAACAGACACCAAGATGATGATGATGATGATGATG 1414
1737 ACGGCACTTACTCCCAAGCATGACAGACCGCGAGTAAAGATTCAGATGATGATGATGATGATG 1796
1415 GTCTGTGTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1474
1797 TTGAT 1856
1475 ATTGCTTACCCCAAGTCAAGTGT-----GGTGGCCGCAACTTAAGCAACAG 1524
1857 ACAGACCGGAGCATGCGAGGTGATGACTAAAGTATTCGCGCTCGACAGCAAAAC 1916
1525 CAAACTG--TCATGCGATTTGCTTCAAAAGTGGCTTCAAGATGATGATGATGATGATGATGATG 1582
1917 CAACTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1976
1583 GAGAGCTGTGAGGAGTGAATCCCTGGAAGCTGATGATGATGATGATGATGATGATGATGATGAT 1642
1977 GAGCTCCCTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2036
1643 ACCATGACATGACAGCTGGGCGAGTCAATGCAAGATTTGTTCCAGATCAATGAAG 1702
2037 GCCATTCACCGAATAATTAAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATG 2096
1703 GAGATGAC--CGCTGATGCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1759
2097 AGAGTCTTTCGCTTACTTCTCCACGCTTAACAGACATTAAGAGGCGCAAGATGATGATGATG 2156
1760 ATGGCTTCAAAAGTGTGCTGCAAGCGCTCTGAGGCTTGAAGTGAATGCTGATGATGATGATG 1819
2157 AGCAGATGTCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2216
1820 TGCCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1879
2217 TGCCAGAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2276
1880 TGAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1939

Wed May 12 09:50:20 2004

us-10-043-774b-1.rng

Page 20

```
Db      2277 TAAACAGCGAGTGAACACCCCTTAAGAGCAGGCTCGACGAAATTTACAAATCAGCTGCA 2336
QY      1940 CTAGACTTAACGTAATTTGTGTGTAAGAAAAGAGTGAACACAGATTTTGTGCTGAGTGTG 1999
Db      2337 AACAGAGGGGCTGTGCGATGACATTTATTAATGTATCCAAAGCGCATTAATTCGCTACT 2396
QY      2000 GAGGAAGACTTCAGATCCACTTCTGTGAAACAGTTATGTGTAGAGGTTACGAGCCAG 2059
Db      2397 TTACTGGGCATCGCAACCCAGTTCGGGCACTGTAGTGAAGAGTTTATTAACCTTGCCAG 2456
QY      2060 AATGGTATGACTTTTATCGGTGAGCCAGGCTGTGAGAGTGTAGTGTTCCTCCACAC 2119
Db      2457 AGCGCTAGACTTCTTCTCTAGTGTGCCAGGCTGTGGATAGGAAGTGTGTGCTACCA 2516
QY      2120 ATTACATGTCTATGACCAACAGCGGCTGAAGCCAGACCATACAGCGCTGACCT 2179
Db      2517 GCTACAAATGTTATTTCTGACCAACATGGGACTAAAGCCGATTAAGCTGAGATGCTCTCT 2576
QY      2180 ACAAGCTGTGCGACATCTATACACTGCGCAGGTGTCAATCGTGTCTGCTCTGCGC 2239
Db      2577 ATTAAGATGACCATATGTACTACAAATTAACAGCGGACCATACAGGTCCCGCTGTGCGC 2636
QY      2240 AGTAAGCCGACAGCTGCGCTTTCTTGTGGCCAGAGTATTCACAGAGCCAAATCTGT 2299
Db      2637 ATTAGGCCCAAAATGTGCTTCTCTGAGCCGAATCCATTAATCGTGGCCTTCAGCAG 2696
QY      2300 CACTGTCAAACGCGCTTACTACGCTTA 2328
Db      2697 GACTGCAGAAATCAATGTACTTTTGTAA 2725
```

Search completed: May 8, 2004, 11:37:40
Job time : 936 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 09:23:26 ; Search time 9065 Seconds

(without alignments)
1131.008 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328
Sequence: 1 atgattcttgggtgaacac.....accgccttactactctctaa 2328

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 segs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

GenBml:
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_hlv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2328	100.0	2328	9	AF264004	AF264004 Homo sapi
2	2316.2	99.5	3421	9	AF387507	AF387507 Homo sapi
3	2313	99.4	3362	9	AF104260	AF104260 Homo sapi
4	2313	99.4	3394	9	BC028581	BC028581 Homo sapi
5	2200.2	94.5	2579	6	AX747944	AX747944 Sequence
6	2200.2	94.5	2579	9	AK093133	AK093133 Homo sapi
7	1684.6	72.4	3910	10	AF438405	AF438405 Mus muscu
8	1684.6	72.4	4064	10	AF438405	AF438405 Mus muscu
9	1013.2	43.5	3273	5	AF363636	AF363636 Danto rer
10	913.2	39.2	3504	9	AB079368	AB079368 Homo sapi
11	680	28.2	3006	9	BC031060	BC031060 Homo sapi
12	680	28.2	3138	6	AX748151	AX748151 Sequence
13	680	28.2	3138	9	AK093475	AK093475 Homo sapi
14	616.8	26.5	2456	6	AX834401	AX834401 Sequence
15	616.8	26.5	2456	9	AK096837	AK096837 Homo sapi
16	568.6	24.4	3057	3	AY014899	AY014899 Strongylo
17	503.2	21.6	3458	3	AK116906	AK116906 Ciona int
18	499.4	21.5	2605	3	AY014800	AY014800 Strongylo
19	460.6	18.8	4913	10	AB032605	AB032605 Mus muscu
20	439.8	18.9	3545	9	AB079367	AB079367 Homo sapi
21	439.8	18.9	3610	9	BC025995	BC025995 Homo sapi
22	435.2	18.7	3620	6	AX714218	AX714218 Sequence
23	435.2	18.7	3620	6	AX056418	AX056418 Homo sapi
24	435.2	18.7	3620	9	AX247684	AX247684 Sequence
25	381	16.4	3649	6	AF285586	AF285586 Mus muscu
26	381	16.4	3659	10	AF014901	AF014901 Strongylo
27	379.4	16.3	1723	3	BC049495	BC049495 Danto rer
28	330.2	14.2	1455	5	AX879653	AX879653 Sequence
29	330	14.2	2272	6	BD157952	BD157952 Primer fo
30	330	14.2	2272	6	BD157952	BD157952 Primer fo
31	330	14.2	2272	9	AK027497	AK027497 Homo sapi
32	310.6	13.3	2761	3	AF145680	AF145680 Drosophi
33	286.4	12.3	300	6	BD212731	BD212731 Novel hum
34	261.8	11.2	2982	3	AF104109	AF104109 Drosophi
35	257	11.0	3047	3	AF104354	AF104354 Drosophi
36	255.4	11.0	3054	3	BT011138	BT011138 Drosophi
37	237.2	10.2	1733	6	AX876238	AX876238 Sequence
38	237.2	10.2	1733	6	BD156024	BD156024 Primer fo
39	237.2	10.2	1733	9	AK001213	AK001213 Homo sapi
40	222.4	9.6	63578	9	AC127071	AC127071 Homo sapi
41	220.8	9.5	193697	2	AC025837	AC025837 Homo sapi
42	203.4	8.7	61995	2	AC090565	AC090565 Homo sapi
43	184	7.9	7781	3	DMDNARPL9	X94613 D.melanogas
44	184	7.9	37980	3	AC005116	AC005116 Drosophi
45	184	7.9	78507	2	AC019704	AC019704 Drosophi

ALIGNMENTS

RESULT 1
AF264004
LOCUS AF264004
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF264004
VERSION AF264004.1 GI:15216446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2328)
Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N.,
Weller,K.P. and Hoffman,R.
Human CD34+ stem cells express the hiwi gene, a human homologue of

Db	1561	CAGATGAAC	TGCAAGATGGGAGGAGAGCTGTGAGGGGTGACATTC	CCCCCTGAAC	TCGTGT	1520
Qy	1621	ATGATCGT	TGGCATGATTCATTTACCATGACATGACAGCTGGGCGGAGGTCAATG	CGACGA		1680
Db	1621	ATGATCGT	TGGCATGATTCATTTACCATGACATGACAGCTGGGCGGAGGTCAATG	CGACGA		1680
Qy	1681	TTGTGTGC	CAGCATCAATGAAGGAGTACCCGCTGGTTCCTCAGCTGCAATTTTCAGGAT			1740
Db	1681	TTGTGTGC	CAGCATCAATGAAGGAGTACCCGCTGGTTCCTCAGCTGCAATTTTCAGGAT			1740
Qy	1741	AGAGGAC	CGAGAGCTGTGATGGGCTCAAACTCTGCTCGAAGGGCTCTTGAGGGCTTTGG			1800
Db	1741	AGAGGAC	CGAGAGCTGTGATGGGCTCAAACTCTGCTCGAAGGGCTCTTGAGGGCTTTGG			1800
Qy	1801	AATAGCTG	CAATGATGATCATGCGCCAGCGGATCATCGTACCGCGATGGCGTAGAGAC			1860
Db	1801	AATAGCTG	CAATGATGATCATGCGCCAGCGGATCATCGTACCGCGATGGCGTAGAGAC			1860
Qy	1861	GGCCAGCT	GAAACACATGCTGATGACACTGCAAGTGCACAACTTTTGGATTGCTTAAATCC			1920
Db	1861	GGCCAGCT	GAAACACATGCTGATGACACTGCAAGTGCACAACTTTTGGATTGCTTAAATCC			1920
Qy	1921	ATTGTTAA	GAGTTACACCTTACACTACCGTAAATTTGTGTGAAGAAAGTGAACCC			1980
Db	1921	ATTGTTAA	GAGTTACACCTTACACTACCGTAAATTTGTGTGAAGAAAGTGAACCC			1980
Qy	1981	AGATTTT	TGCTCACTGTGAGGAAGACTTGAGATCCACTTCCGTGAAACAGTTATGAT			2040
Db	1981	AGATTTT	TGCTCACTGTGAGGAAGACTTGAGATCCACTTCCGTGAAACAGTTATGAT			2040
Qy	2041	GTAGAAGT	TACCAAGCCGAATGTTATGACTTTTATTCGTGACCGCAGGCTGTGAAGT			2100
Db	2041	GTAGAAGT	TACCAAGCCGAATGTTATGACTTTTATTCGTGACCGCAGGCTGTGAAGT			2100
Qy	2101	GGTAGTGT	TTCTTCCCAACATTAATCTTGTGACCAACAGGGGCTGAAGCCAGAC			2160
Db	2101	GGTAGTGT	TTCTTCCCAACATTAATCTTGTGACCAACAGGGGCTGAAGCCAGAC			2160
Qy	2161	CACATACAG	CGGTTGACCTTACAGCTGTGACACATCTATTAACAATGCGCCAGGTGTCAAT			2220
Db	2161	CACATACAG	CGGTTGACCTTACCTTACCTTACCAAGCTGTGCCACATCTTATTAACAATGCGCCAGGTGTCAAT			2220
Qy	2221	CGTGTCTC	TGCTCCTTGGCAGTACGCCCAACAAGCTGGCTTTCTTGTGGCCAGAGTAT			2280
Db	2221	CGTGTCTC	TGCTCCTTGGCAGTACGCCCAACAAGCTGGCTTTCTTGTGGCCAGAGTAT			2280
Qy	2281	CACAGAGAG	CCCAATCTGTCACTGTCAAAACCGCTTTACTACTCTTAA			2328
Db	2281	CACAGAGAG	CCCAATCTGTCACTGTCAAAACCGCTTTACTACTCTTAA			2328
RESULT 2						
AF387507						
LOCUS						
DEFINITION Homo sapiens PIWI protein mRNA, complete cds.						
ACCESSION AF387507						
VERSION AF387507.1 GI:14579644						
KEYWORDS						
SOURCE						
ORGANISM						
Homo sapiens (human)						
Home sapiens						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.						
REFERENCE						
AUTHORS						
TITLE						
Cloning and identification of human piwi protein related to testis						
development						
Unpublished						
2 (bases 1 to 3421)						
Sha,J.H., Li,J.M. and Zhou,Z.M.						
Direct Submission						
Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing						
Medical University, 140 Han Zhong Road, Nanjing Medical University						
Jiangsu 210029, China						

FEATURES	SOURCE	Location/Qualifiers
		1..3421
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/tissue_type="testis"
		/dev_stage="adult"
		64..2649
		/note="higher expression in adult testis than 6 month old embryo testis"
		embryo testis=1
		/product="PIWI protein"
		/protein_id="AAK69348.1"
		/db_xref="GI:14579645"
		/translation="MTGRARARARARGAGGTAGTGVSTASQGPVYIQRPDPPEAGELFRSGRORGATGATAGTAKSQGLISAGFOELSGRGRRRRPHLGAVTRNDLVHKEKSTGSSGIIIVRLSTNNHRLTSPQWALYOHIDVNPALRSLALFQHEIDIGKKEHAPGCTGLFLPKRQCKVTEPVSSTKRGEDVRITLTLENLSPSPLOQFYNIIPRRHLKTMNMQOIGRNTINRNDPIDIPSHLVYMPGTTSLIOYENSIMCTDVSHVLRSEBVDFEMFNFYHQTBEHKEFEOVSKELGLVNLTKNNKTIRVDDIMDQNSKTPFKADGSEVSELYEYRQYNQFETDLQPVAVSQPKRRRGGLYLPENMLPELCITGLTDRKRDENVMKDLAVHRLPBPOROREVGLIDYIHKNDVQSELHDWGSFDSNLSFGRKRIOTEKIHQGGKTFDNPQFADMSKETRAPLISVPLNNMLIYTRYRYEALNSLIONFKYTPAMGMOMRKAIMTEVDRDTRMAYRIVQOKTAPATQIYCLSNRKKKIDYDLKXILCTDCTPSCQCVARVATLQKQYTAALATKALOMKMGSELRWYDIPLEKLVMLVGIDCYHDMDTGRSLAGFYASINBGTTRFSPRCIPDRBOELVYDGLKVCGLALVLRANSNENYSPRIIVYRBDVSGQLKTLVNVVPOPLDCKSKTGRNRYRLVYIVLKRNVTRFAQSGRLONPLPGTVIIDEVTRPEVYDFPVSQAVRSGVSFTNNVYINSLGKPDHILQRLTYKLCHIIYNNPVGIVRPAPQVAKHKAFLVGQSTHREPNLSLSRLYYL"
		3354..3359
		PolyA_signal
		ORIGIN
		Query Match 99.5%; Score 2316.2; DB 9; Length 3421;
		Best Local Similarity 99.9%; Pred. No. 0;
		Matches 2318; Conservative 0; Mismatches 3; Indels 0; Gaps 0
QY	8	TTGGTGTAAACAACAAGGCAAACTTAGACATGTTAAAGATCAAAAACAGGTTCTTGAG 67
DB	329	TTCGTGTAAATACAGGCAAACTTAGACATGTTAAAGATCAAAAACAGGTTCTTGAG 388
QY	68	GCATTATAGTAAAGTTAAGCACTAACATTTCCGGCTACATCCCGTCCCACTGGGCGCT 127
DB	389	GCATTATAGTAAAGTTAAGCACTAACATTTCCGGCTACATCCCGTCCCACTGGGCGCT 448
QY	128	TATATCATATCACTTGAATACCACTGATGGAAGCCAGAGACATCCGTTGAGCTC 187
DB	449	TATATCATATCACTTGAATACCACTGATGGAAGCCAGAGACATCCGTTGAGCTC 508
QY	188	TTCTTTTTCACACACGAGATCTAATTGGAAGTGCATGCTTTTGATGGAACGATATTAT 247
DB	509	TTCTTTTTCACACGAGATCTAATTGGAAGTGCATGCTTTTGATGGAACGATATTAT 568
QY	248	TTTTACCTTAAAGATCTACAGCAAAAGGTTACTGAAAGTTTATAGTAAAGCCGGAATGAG 307
DB	569	TTTTACCTTAAAGATCTACAGCAAAAGGTTACTGAAAGTTTATAGTAAAGCCGGAATGAG 628
QY	308	AGGATGTGAGATTAAGATCACTTTTAACAAATGAATCTTCAACCTTCAATCAACCACTGTT 367
DB	629	AGGATGTGAGATTAAGATCACTTTTAACAAATGAATCTTCAACCTTCAATCAACCACTGTT 688
QY	368	TGCAGTTCATATATAATTATTTTTCAGAGGCTTTGAAAATCATGAAATTGGCAAAATG 427
DB	689	TGCAGTTCATATATAATTATTTTTCAGAGGCTTTGAAAATCATGAAATTGGCAAAATG 748
QY	428	GACGAAATATATATACCAACCAATGACCAATGATATTCACAGTCAAGGTGTGATTT 487
DB	749	GACGAAATATATATACCAACCAATGACCAATGATATTCACAGTCAAGGTGTGATTT 808
QY	488	GGCCTGGCTTCACTACTTCATCTTCACTGATGAAAACAGCATCATGCTCTGCACTGAG 547
DB	809	GGCCTGGCTTCACTACTTCATCTTCACTGATGAAAACAGCATCATGCTCTGCACTGAG 868

QY 548 TTAGCCATAAAGTCTTGAAGTGAAGCTTTTGAATTCATGTTCACTTTATCATC 607
 Db 869 TTACCCATAAAGTCTTGAAGTGAAGCTTTTGAATTCATGTTCACTTTATCATC 928
 QY 608 AGACAGAGAAGACATTAATTTCAAGAACAGTCTTCAAGAACTAATAGTTAGTTGTC 667
 Db 929 AGACAGAGAAGACATTAATTTCAAGAACAGTCTTCAAGAACTAATAGTTAGTTGTC 988
 QY 668 TTACCAAGTATACATTAAGACATACAGAGTGAATGATATTGACTGGAGCCAAATCCCA 727
 Db 989 TTACCAAGTATACATTAAGACATACAGAGTGAATGATATTGACTGGAGCCAAATCCCA 1048
 QY 728 AGACACCTTTAAGAAAGCCGCTCTGAGTCAAGCTTCTTGAATATCTACAGGAAGC 787
 Db 1049 AGACACCTTTAAGAAAGCCGCTCTGAGTCAAGCTTCTTGAATATCTACAGGAAGC 1108
 QY 788 AATACACCAAGAGATCACCGACTTGAAGACCTGTCTTGTGTCAGCCCAAGAGAA 847
 Db 1109 AATACACCAAGAGATCACCGACTTGAAGACCTGTCTTGTGTCAGCCCAAGAGAA 1168
 QY 848 GAGCGGGGCTGGGGGGGCACTGCGAGGGCTGCTCATCTGAGCTCTGCTATC 907
 Db 1169 GAGCGGGGCTGGGGGGGCACTGCGAGGGCTGCTCATCTGAGCTCTGCTATC 1228
 QY 908 TTACAGGTCTAATGATTAATGCGTAAATGATTTTAACTGTAGTAAAGACTTACCGCTTC 967
 Db 1229 TTACAGGTCTAATGATTAATGCGTAAATGATTTTAACTGTAGTAAAGACTTACCGCTTC 1288
 QY 968 AATACAGCTAATCCAGAGAAAGGAGGAGTGAAGTGGGAGCACTGATGATTAATC 1027
 Db 1289 AATACAGCTAATCCAGAGAAAGGAGGAGTGAAGTGGGAGCACTGATGATTAATC 1348
 QY 1028 AATAAAAGATATGTTCAAGAGGAGCTTGAAGCTGGGGTTGAGCTTTGATTCACAT 1087
 Db 1349 AATAAAAGATATGTTCAAGAGGAGCTTGAAGCTGGGGTTGAGCTTTGATTCACAT 1408
 QY 1088 TACGTCTCTCTCAGAGAAATTTTGCAGAAAGAAAGATCACCAAGGTGGAAGAAATCAT 1147
 Db 1409 TACGTCTCTCTCAGAGAAATTTTGCAGAAAGAAAGATCACCAAGGTGGAAGAAATCAT 1468
 QY 1148 TTGATTAATCAATCCAAATTTGCAAGTTGCTCAAGAAAGAAAGAGTGCACATTAATTA 1207
 Db 1469 TTGATTAATCAATCCAAATTTGCAAGTTGCTCAAGAAAGAAAGAGTGCACATTAATTA 1528
 QY 1208 GTGTTAAGCCACTAGATTAATCTGCTGTGATCTATAGCCGGAAGAAATTAAGAACAGCA 1267
 Db 1529 GTGTTAAGCCACTAGATTAATCTGCTGTGATCTATAGCCGGAAGAAATTAAGAACAGCA 1588
 QY 1268 ATTCATGATCAAAATCTATTTAAAGTTACACAGCCATGGGATGCAATGAGAAAG 1327
 Db 1589 ATTCATGATCAAAATCTATTTAAAGTTACACAGCCATGGGATGCAATGAGAAAG 1648
 QY 1328 CATTAATGATGAGTGAATGACAGAACTGAAGCTTAAAGTCTTAAAGCAAAAG 1387
 Db 1649 CATTAATGATGAGTGAATGACAGAACTGAAGCTTAAAGTCTTAAAGCAAAAG 1708
 QY 1388 TCACAGCAGACCCAGATAGTGTCTGCTGTGCTCAAGTAAATCGAGAGCAAAATACG 1447
 Db 1709 TCACAGCAGACCCAGATAGTGTCTGCTGTGCTCAAGTAAATCGAGAGCAAAATACG 1768
 QY 1448 ATGCTATTAAATAATACCTGTGTACAGATTGCTCCCAAGTCAAGTGTGTGTGCC 1507
 Db 1769 ATGCTATTAAATAATACCTGTGTGTACAGATTGCTCCCAAGTCAAGTGTGTGTGCC 1828
 QY 1508 GAACTTGAAGCAAAAGCAAGCAATGTCATGAGCCATGCTCAAAAGTCCCTACAGATGA 1567
 Db 1829 GAACTTGAAGCAAAAGCAAGCAATGTCATGAGCCATGCTCAAAAGTCCCTACAGATGA 1888
 QY 1568 ACTGCAAGTGGAGAGAGCTCTGAGAGGTGAGACATCCCTGTAAGCTCTGATGATCG 1627
 Db 1889 ACTGCAAGTGGAGAGAGCTCTGAGAGGTGAGACATCCCTGTAAGCTCTGATGATCG 1948
 QY 1628 TTGGCATGATTTGTTACATGACATGACAGCTGGGGGAGAGTCAATGCCAGAGATTTGTTG 1687

Db 1949 TTGGCATGATTTGTTACATGACATGACAGCTGGGGGAGGTCAATGCAAGATTTGTTG 2008
 QY 1688 CCAGCATCAATGAAGAGATGACCCGCTGTCTTCAAGCTGATATTTTCAAGATGAGAGAC 1747
 Db 2009 CCAGCATCAATGAAGAGATGACCCGCTGTCTTCAAGCTGATATTTTCAAGATGAGAGAC 2068
 QY 1748 AGAGCTGTATGATGGGCTCAAGAGTGTGCTGCAAGCGGCTCTGAGGGCTTGAATAGCT 1807
 Db 2069 AGAGCTGTATGATGGGCTCAAGAGTGTGCTGCAAGCGGCTCTGAGGGCTTGAATAGCT 2128
 QY 1808 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867
 Db 2129 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2188
 QY 1868 TGAATACTGGTGAATACGAGAGTGCAGAGTCTTTTGAATGCTTAAATCAATGCTA 1927
 Db 2189 TGAATACTGGTGAATACGAGAGTGCAGAGTCTTTTGAATGCTTAAATCAATGCTA 2248
 QY 1928 GAGTTTACACCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1987
 Db 2249 GAGTTTACACCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2308
 QY 1988 TTGCTCAGCTGAG 2047
 Db 2309 TTGCTCAGCTGAG 2368
 QY 2048 TTACCAAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2107
 Db 2369 TTACCAAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2428
 QY 2108 TTTTCCCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2167
 Db 2429 TTTTCCCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2488
 QY 2168 AGCGCTTGAACCTTGAAGCTGTCACATCTATTAATCAATGATGATGATGATGATGATGAT 2227
 Db 2489 AGCGCTTGAACCTTGAAGCTGTCACATCTATTAATCAATGATGATGATGATGATGATGAT 2548
 QY 2228 CTGCTCTTGAAG 2287
 Db 2549 CTGCTCTTGAAG 2608
 QY 2288 AGCCAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
 Db 2609 AGCCAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2649
 RESULT 3
 AF104260 3362 bp mRNA linear PRI 30-APR-2003
 LOCUS AF104260
 DEFINITION Homo sapiens HIWI mRNA, complete cds.
 VERSION AF104260
 KEYWORDS AF104260.2 GI:18098557
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 3362)
 Cox D.N., Chao A., Baker J., Chang L., Qiao D. and Lin H.
 A novel class of evolutionarily conserved genes defined by piwi are
 essential for stem cell self-renewal
 Genes Dev. 12 (23), 3715-3727 (1998)
 JOURNAL
 MEDLINE 99069219
 PUBMED 9851978
 REFERENCE 2 (bases 1 to 3362)
 Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J. and Lin H.
 Molecular characterization of hiwi, a human member of the piwi gene
 family whose overexpression is correlated to seminomas
 Oncogene 21 (25), 3988-3999 (2002)
 JOURNAL
 MEDLINE 22035724
 PUBMED 12037681

REFERENCE 3 (bases 1 to 3362)
 AUTHORS Lin, H.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1998) Cell Biology, Duke University Medical Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710, USA

REFERENCE 4 (bases 1 to 3362)
 AUTHORS Qiao, D., Zeeman, A.-M., Deng, W., Looijenga, J. H. J., and Lin, H.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-2002) Cell Biology, Duke University, 412 Nanaline Duke Building, DUMC, Durham, NC 27710, USA

REMARK Nucleotide sequence updated by submitter
 COMMENT On Jan 9, 2002 this sequence version replaced gi:4038412.
 FEATURES
 location/Qualifiers
 1..3362
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12q24.33"
 /issue_type="testis"
 57..2642
 /note="member of the PIWI family"
 /product="Piwi"
 /protein_id="AAC97371.2"
 /db_xref="GI:18098558"

CDS

translation="MTGARARARARAGGTAAQVGTASQCPGYIQPRPPPPAG
 ELPRGRGRGTAGTAKSGOGLISAGPDELISLGRGRDRPDLGVNTQNDHYE
 SKTSGSGLIVLSTHPELSTHPELSTHPELSTHPELSTHPELSTHPELSTH
 HAFDSTLPLKRLQOKTVEPSKTRNVEDRITITITLPELSTHPELSTHPEL
 LKLTMLQOIGRNTYNPDPIDIPSHRVLVPGFTTSILQYENSIMCTVSHVES
 ETVLDFEMFPHQTEHNFQVSKELIGLVLRKYNKYRVDDIDMDQVSKTEK
 ADGGSVLELYRKOYNOEITDLKQPVLSQPKRRRGGTLPBPAMLPELCTVGL
 TDKRNDFWVKMDLAVHRLPEGRORSEVGLIDYIHNNDVQSELMDWGLSPSNIL
 SPGRILLOTETKHOGKTFDNPQPADMSKETRGAPLISVPELDNMLIYTRNVEA
 NSLIONLKTPTPMGMOMKRAIMTEVDRTAAYLRVLOKVTADTOIVCLSNRQ
 KYDAIKKLTCDPTSPCCVARTIGKQTMALATKIALOMKMGELMRNDPIPK
 LVMIVGIDYHDMTARSRISAGFVASINSGTMRSPRTFDRQDELIDGKVLQAA
 LRANSCNEMPSRIIVRDVGGDLKTVNVEVPPDLCKSIGRINRGLIVY
 KKRVTREFAGSGRLONPLPGVLDVETRPEDVPEVQAVRSQSVSTHNAVY
 DNSGLKPHIQLRTYKLCIYINWGVIRVPACQYAKLAFLVGSISHRPNLSLN
 RLIVL"

ORIGIN

Query Match 99.4%; Score 2313; DB 9; Length 3362;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACACAGGCGAGACTAGACCATGTTAAAGATCAAAAACAGTTCTTCAAG 67
 DB 322 TTGGTGTGAATACAGGCGAGACTAGACCATGTTAAAGATCAAAAACAGTTCTTCAAG 381
 QY 68 GCATTATAGTAAAGTTAAGCACTAACCATTTCCGGCTGACATCCGTCCTCCAGTGGGCTT 127
 DB 382 GCATTATAGTAAAGTTAAGCACTAACCATTTCCGGCTGACATCCGTCCTCCAGTGGGCTT 441
 QY 128 TATATAGATACATGATGATTAACCCCTAGATGGAAGCGAGAAAGACTCCGCTTCAAGCTC 187
 DB 442 TATATAGATACATGATGATTAACCCCTAGATGGAAGCGAGAAAGACTCCGCTTCAAGCTC 501
 QY 188 TTCTTTTCAACAGAGATCTAATTTGAAAGTGTGATGCTTTTGAATGGAACGATATTAT 247
 DB 502 TTCTTTTCAACAGAGATCTAATTTGAAAGTGTGATGCTTTTGAATGGAACGATATTAT 561
 QY 248 TTTTACCTTAAAGACTACAGCAAAAGTTACTGAAGTTTATAGTAAAGACCGGAAATGAG 307
 DB 562 TTTTACCTTAAAGACTACAGCAAAAGTTACTGAAGTTTATAGTAAAGACCGGAAATGAG 621
 QY 308 AGGATGTAGATACATGATCACTTTAAACAATGAACTTCCACCTACATCAACCACTTGT 367
 DB 622 AGGATGTAGATACATGATCACTTTAAACAATGAACTTCCACCTACATCAACCACTTGT 681

QY 368 TCGAGTCTATATATATATATTTTCAAGAGGCTTTGAAAATCATGATTTGCAACAAATTTG 427
 DB 682 TCGAGTCTATATATATATATTTTCAAGAGGCTTTGAAAATCATGATTTGCAACAAATTTG 741
 QY 428 GACGAATTTATATACCAAAATGACCCATTTGATTTCCAGTACAGAGTGGTATTT 487
 DB 742 GACGAATTTATATACCAAAATGACCCATTTGATTTCCAGTACAGAGTGGTATTT 801
 QY 488 GGCCTGGCTTCACTATCTTCATCTTCAGTATGAAAACAGCATCTCTTCCAGTACG 547
 DB 802 GGCCTGGCTTCACTATCTTCATCTTCAGTATGAAAACAGCATCTCTTCCAGTACG 861
 QY 548 TTAGCCATTAAGTCTTCCAGATGAGACTGTTTGGATTTCAATTTTATCATTC 607
 DB 862 TTAGCCATTAAGTCTTCCAGATGAGACTGTTTGGATTTCAATTTTATCATTC 921
 QY 608 AGACAGAGACATTAATTTCAAGAACCAAGTTTCCAAAGACATTAAGTATTGTTTC 667
 DB 922 AGACAGAGACATTAATTTCAAGAACCAAGTTTCCAAAGACATTAAGTATTGTTTC 981
 QY 668 TTACCAAGTATTAACATTAAGACATACAGTGTGATTTTGACTGGGACCAAGTCCCA 727
 DB 982 TTACCAAGTATTAACATTAAGACATACAGTGTGATTTTGACTGGGACCAAGTCCCA 1041
 QY 728 AGAGCACTTTAAGAAAGCCGACGCGCTGAGTCAAGTCAAGTCTTCTAGATCTACAGGAAC 787
 DB 1042 AGAGCACTTTAAGAAAGCCGACGCGCTGAGTCAAGTCTTCTAGATCTACAGGAAC 1101
 QY 788 AATACACACAGAGATCACCGACTTGAAGACGCTGCTTGTGTCAAGCCCAAGAGA 847
 DB 1102 AATACACACAGAGATCACCGACTTGAAGACGCTGCTTGTGTCAAGCCCAAGAGA 1161
 QY 848 GGGGGGGCCCTGGGGGGGACATGCGCAGGCGCTGCACTTCTGTAGCTCTGTATC 907
 DB 1162 GGGGGGGCCCTGGGGGGGACATGCGCAGGCGCTGCACTTCTGTAGCTCTGTATC 1221
 QY 908 TTAAGGTCTTAAGTAAATTCGTAATGATTTTAACGATGAAGACTTACGCGTTC 967
 DB 1222 TTAAGGTCTTAAGTAAATTCGTAATGATTTTAACGATGAAGACTTACGCGTTC 1281
 QY 968 ATACAGACATTAACCGAGCAAGGAGGAGTGAAGTGGGAGAGCATGATGATTAATTC 1027
 DB 1282 ATACAGACATTAACCGAGCAAGGAGGAGTGAAGTGGGAGAGCATGATGATTAATTC 1341
 QY 1028 ATAAAAACATTAATGTTCAAGAGGAGCTTGAAGTGGGTTTGAAGCTTATTCACACT 1087
 DB 1342 ATAAAAACATTAATGTTCAAGAGGAGCTTGAAGTGGGTTTGAAGCTTATTCACACT 1401
 QY 1088 TACTGCTCTTCAAGAGAAATTTTGGCAACAGAAAAGTTTCAACCAAGTGGAAAAACAT 1147
 DB 1402 TACTGCTCTTCAAGAGAAATTTTGGCAACAGAAAAGTTTCAACCAAGTGGAAAAACAT 1461
 QY 1148 TTGATTAACATTCACAAATTTGAGATTTGTCACAAAAGAAACAAAGGGTGCACATTAATTA 1207
 DB 1462 TTGATTAACATTCACAAATTTGAGATTTGTCACAAAAGAAACAAAGGGTGCACATTAATTA 1521
 QY 1208 GTGTTAAGCACTAGATTAATCTGCTGTGATCTTATACGCAAGAAATTAAGAAAGGCGCA 1267
 DB 1522 GTGTTAAGCACTAGATTAATCTGCTGTGATCTTATACGCAAGAAATTAAGAAAGGCGCA 1581
 QY 1268 ATTCAATGATTAACAAATCTAATTAAGTACACAGGATGGGATGCAATGAGAAAG 1327
 DB 1582 ATTCAATGATTAACAAATCTAATTAAGTACACAGGATGGGATGCAATGAGAAAG 1641
 QY 1328 CAATTAATGATTAAGTGAATGACAGAACTGAAGCTTCTTAAGAGTTTCAAGAAAG 1387
 DB 1642 CAATTAATGATTAAGTGAATGACAGAACTGAAGCTTCTTAAGAGTTTCAAGAAAG 1701
 QY 1388 TCAAGAGACACCCCAATGATGTTGTCTGTCTGTCTCAAGTAACTGGAAGAGCAAAATACG 1447
 DB 1702 TCAAGAGACACCCCAATGATGTTGTCTGTCTGTCTCAAGTAACTGGAAGAGCAAAATACG 1761
 QY 1448 ATGCTATTAATAAAATACCTGTGTAAGATTCCTACCCCAAGTCACTGTGTGTGCGCC 1507

SKTSGSGLIVLSTNHPRLTSRPMOALYOHYDINPLMEARLSALLFOHEDLIGKC
HAFQITLFLPKRLOQKVEKSRKNEDEDRITITLNLNPEPSPCLQFYNILFER
LTKMNDQIGRNTNRPDIPIDIPSHLVITPFTTSLIOENSIMLCTDYSKVLRS
ETVLDPMFNPHQTEEHKFOQVSKEELIGLVLTINKTKTRVDIDMDONPKSTPEK
ADGSEVSLFLEYRKOYNQISITDKQPVLSQPKRRGGGTRPFPALIPELCTLTG
TDKRMNDENWMDLAVHRLTEPQORQEVGRILIDYIKNDVQSELSDWGLSPSNL
SFSGRILLOTEKIHOGKTFDYNQPOPADMSKETRGAFLISVPLNMLLIYTRNVEA
NSLIONLFRTPAMGOMKKAIMEVDRTAVLRLOQKVTADTQIVCLLSNRD
KYDAIKYPCDPCPTSPQCVARLTGQOMTALATKIALQMCNKGELMRVYDPIK
LMTVIGCTHEDMTGRSISAGFASISIEGTWRPSCITPDROEGLVDGLKVLQAA
LWAMSCEHYESRLIVRVDGSDQLTLVNEYRPOGLDKSISGSGVSRHVLIVY
KRVNTRFQSGRLQNPDPETVIDVTEPMDYDFIVQASGSVSETHNLYN
DMSGKPHIQLRLTYKLOHIYMWGVIRVAPCQYAKRLAFLVQSISHRPNLSLN
RLYL."

misc_feature
900..1310
/note="PAZ; Region: PAZ domain. This domain is named PAZ
after the proteins Pw1, Argonaut and Zwille. This domain
is found in two families of proteins that are involved in
post-transcriptional gene silencing. These are the Pw1
family and the Dicer family, that includes the Carpel
factory protein. The function of the domains is unknown
but has been suggested to mediate complex formation
between proteins of the Pw1 and Dicer families by
hetero-dimerisation"
/db_xref="CDD:Pfam02170"
1728..2609
/note="Pw1; Region: Pw1 domain. This domain is found in
the protein Pw1 and its relatives. The function of this
domain is unknown"
/db_xref="CDD:Pfam02171"

ORIGIN
Query Match 99.4%; Score 2313; DB 9; Length 3394;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

8 TTGGTGTGAACACAGGCGAAGTCTAGACATGTTAAGATCAAAAAGAGTCTTCAG 67
334 TTGGTGTGAATAAAGCGAGAACCTAGACATGTTAAGATCAAAAAGAGTCTTCAG 393
68 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGGTGACATCCGCTCCCAAGTGGCCT 127
394 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGGTGACATCCGCTCCCAAGTGGCCT 453
128 TATATCAGTATCACTGATTAACCCACTGATGGAAGGCAAGAACCTCCGTTGAGTGC 187
454 TATATCAGTATCACTGATTAACCCACTGATGGAAGGCAAGAACCTCCGTTGAGTGC 513
188 TTCTTTTCAACAGAGATCTAATGGAAGTGTCACTTTTGTGGAAGCAATATAT 247
514 TTCTTTTCAACAGAGATCTAATGGAAGTGTCACTTTTGTGGAAGCAATATAT 573
248 TTTTACCTAAAGCTACAGCAAAAGGTTACTGAAGTTTGTAGTAAAGCCCGGAATGGAG 307
574 TTTTACCTAAAGCTACAGCAAAAGGTTACTGAAGTTTGTAGTAAAGCCCGGAATGGAG 633
308 AGGATGAGAGATACAGTACCTTAACAATGAAGTCCACCTACATCACTTGT 367
634 AGGATGAGAGATACAGTACCTTAACAATGAAGTCCACCTACATCACTTGT 693
368 TGCAGTCTATATATATTTTTCAGAGGCTTTGAAAATCAATGAATTTGCAACAAATG 427
694 TGCAGTCTATATATATTTTTCAGAGGCTTTGAAAATCAATGAATTTGCAACAAATG 753
428 GACGAATTTATTAACCCAAATGACCAATGATTTTCAAGTCAAGGTGGTATTT 487
754 GACGAATTTATTAACCCAAATGACCAATGATTTTCAAGTCAAGGTGGTATTT 813
488 GAGCTGAGTCTACTACTTCATCTTCAGATGAAAAAGAGATATGCTGTGCACTAGC 547
814 GAGCTGAGTCTACTACTTCATCTTCAGATGAAAAAGAGATATGCTGTGCACTAGC 873
548 TTAGCATTAAGTCTTGAAGTGAAGATGTTTGGATTTCATGTTCAACTTTATCATC 607

874 TTAGCATTAAGTCTTGAAGTGAAGATGTTTGGATTTCATGTTCAACTTTATCATC 933
608 AGACAGAGAACATAATTTCAAGAACAGATTTCCAAAGAACTAATAGTTTAGTTGTTTC 667
934 AGACAGAGAACATAATTTCAAGAACAGATTTCCAAAGAACTAATAGTTTAGTTGTTTC 993
668 TTACCAAGTTTAACTAATGAACATACAGAGTGAATATTTGACTGAGGACAGAAATCCCA 727
994 TTACCAAGTTTAACTAATGAACATACAGAGTGAATATTTGACTGAGGACAGAAATCCCA 1053
728 AGAGCACTTTAAGAAAGCGAGCTGAGTCAAGTCACTTCTTGAATATCTACAGAAAGC 787
1054 AGAGCACTTTAAGAAAGCGAGCTGAGTCAAGTCACTTCTTGAATATCTACAGAAAGC 1113
788 AATACAAACAAAGATACCCGACTTGAAGCAGCTCTTGTGTCAGCCAGCCCAAGAGAA 847
1114 AATACAAACAAAGATACCCGACTTGAAGCAGCTCTTGTGTCAGCCAGCCCAAGAGAA 1173
848 GGGGGGGCCCTGGGGGGGACCTGCGAGGGCTGCATGCTCATTCCTGAGCTGTCTATC 907
1174 GGGGGGGCCCTGGGGGGGACCTGCGAGGGCTGCATGCTCATTCCTGAGCTGTCTATC 1233
908 TTACAGGCTTAAGTAAATGCGTATGATTTTAACTGATGAAAGACTTGAACCGCTTC 967
1234 TTACAGGCTTAAGTAAATGCGTATGATTTTAACTGATGAAAGACTTGAACCGCTTC 1293
968 ATACAAAGCTTACTCCAGACCAAGAGAGGTGAAGTGAAGGAGCACTGATGATTAATTC 1027
1294 ATACAAAGCTTACTCCAGACCAAGAGAGGTGAAGTGAAGGAGCACTGATGATTAATTC 1353
1028 ATAAAAAGATATGTTCAAAAGGAGCTTCGAGCTGGGTTTGAAGCTTTGATTCACACT 1087
1354 ATAAAAAGATATGTTCAAAAGGAGCTTCGAGCTGGGTTTGAAGCTTTGATTCACACT 1413
1088 TACTGCTCTTCCAGAGAAATTTTGAACAGAAAGATTCACCAAGTGAAGAAACAT 1147
1414 TACTGCTCTTCCAGAGAAATTTTGAACAGAAAGATTCACCAAGTGAAGAAACAT 1473
1148 TTGATTAACATCCAAATTTGACAGATTGTTCAAGAAACAAAGGTTGACCATTAATTA 1207
1474 TTGATTAACATCCAAATTTGACAGATTGTTCAAGAAACAAAGGTTGACCATTAATTA 1533
1208 GTGTTAAGCCCTAGATTAAGTCTGTTGATCTTAAAGCGAAGAAATATGAGAGAGCA 1267
1534 GTGTTAAGCCCTAGATTAAGTCTGTTGATCTTAAAGCGAAGAAATATGAGAGAGCA 1593
1268 ATTCAATTGATCAAAATCTATTTAAAGTTTACACAGCCATGGGCAATGAGAAAG 1327
1594 ATTCAATTGATCAAAATCTATTTAAAGTTTACACAGCCATGGGCAATGAGAAAG 1653
1328 CAATATATATGAATGATGACAGAACTGAAGCTTATTAAAGTCTTACAGCAAAAG 1387
1654 CAATATATATGAATGATGACAGAACTGAAGCTTATTAAAGTCTTACAGCAAAAG 1713
1388 TCAAGCAGACACCCAGATGTTGTTCTGTTGTCAAGTATTCGGAAGCAAAATAGC 1447
1714 TCAAGCAGACACCCAGATGTTGTTCTGTTGTCAAGTATTCGGAAGCAAAATAGC 1773
1448 ATGCTATTAATAAATACCTGTGTACAGATGTCCTTACCCCAAGTCAAGTGTGTGGGCC 1507
1774 ATGCTATTAATAAATACCTGTGTACAGATGTCCTTACCCCAAGTCAAGTGTGTGGGCC 1833
1508 GAACCTTGAAGCAACAGAACTGTCAAGGCAATGCTTACAAAGATTTGCCCTTACAGATGA 1567
1834 GAACCTTGAAGCAACAGAACTGTCAAGGCAATGCTTACAAAGATTTGCCCTTACAGATGA 1893
1568 ACTGCAAGTGAAGAGAGCTCTGAGAGGTGAGCAATCCCTGAAAGCTGTGATGATG 1627
1894 ACTGCAAGTGAAGAGAGCTCTGAGAGGTGAGCAATCCCTGAAAGCTGTGATGATG 1953
1628 TTGGCATGATTTGTTTCAATGACATGACAGCTGGGGGAGAGTCAATCGAGATTTGTTG 1687

Db 1954 TTGGCATGATGTTTACATGACATGACGCTGGCGGAGGTCATTCGACAGATTGTTG 2013
 QY 1688 CCAGCATCAATGAGAGGATGACCCGCTGGTTCACGCTGCATATTTCAGATAGAGAC 1747
 Db 2014 CCAGCATCAATGAGAGGATGACCCGCTGGTTCACGCTGCATATTTCAGATAGAGAC 2073
 QY 1748 AGGAGCTGTATAGGCTCAAGTCTGCTGCAAGCGGCTCTGAGGCTTGGAAATAGCT 1807
 Db 2074 AGGAGCTGTATAGGCTCAAGTCTGCTGCAAGCGGCTCTGAGGCTTGGAAATAGCT 2133
 QY 1808 GCATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 1867
 Db 2134 GCATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 2193
 QY 1868 TGAATACTGTGTGAATCTGCAAGTGGCCACGTTTGGATTGTCTAAATTCATTGGTA 1927
 Db 2194 TGAATACTGTGTGAATCTGCAAGTGGCCACGTTTGGATTGTCTAAATTCATTGGTA 2253
 QY 1928 GAGGTACACACCTAGACTAAGCTATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAG 1987
 Db 2254 GAGGTACACACCTAGACTAAGCTATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAG 2313
 QY 1988 TTGCTCAGTCTGAG 2047
 Db 2314 TTGCTCAGTCTGAG 2373
 QY 2048 TTACCAAGCAGAGATGCTATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2107
 Db 2374 TTACCAAGCAGAGATGCTATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2433
 QY 2108 TTTCCTCCACACATTAATGCTATGCTATGAGACAGAGGCTGAGAGAGAGAGAGAG 2167
 Db 2434 TTTCCTCCACACATTAATGCTATGCTATGAGACAGAGGCTGAGAGAGAGAGAGAG 2493
 QY 2168 AGGCTTGAACCTGAG 2227
 Db 2494 AGGCTTGAACCTGAG 2553
 QY 2228 CTGCTCTGCTGAG 2287
 Db 2554 CTGCTCTGCTGAG 2613
 QY 2288 AGCCAAATCTGTCACTGTCAAAACCGCTTTTACCTCTTAA 2328
 Db 2614 AGCCAAATCTGTCACTGTCAAAACCGCTTTTACCTCTTAA 2654

RESULT 5
 AX747944 2579 bp mRNA linear PAT 20-JUN-2003
 LOCUS Definition Sequence 1469 from Patent EP1308459.
 AX747944
 VERSION AX747944.1 GI:32132332
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catartini; Homnidae; Homo.

REFERENCE
 AUTHORS
 1 Isegaki, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamehika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and Masuno, Y.

TITLES
 JOURNAL
 Patent: EP 1308459-A 1469 07-MAY-2003;
 Helix Research Institute (JP); Research Association for
 Biotechnology (JP)

FEATURES
 SOURCE
 1. 2579
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 94.5%; Score 2200.2; DB 6; Length 2579;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 8 TTGGTGAACACAGGAG 67
 Db 319 TTGGTGAATACAGGAG 378
 QY 68 GCATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 127
 Db 379 GCATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 438
 QY 128 TATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 187
 Db 439 TATATGATCAATGCGCCAGGATCATGCTGATCCGATGCGGATGAGAGAGCGGACG 498
 QY 188 TTCTTTTCAACAGAGATCTAATGGAAGTGTATGCTTGAATGGAAGATTTAT 247
 Db 499 TTCTTTTCAACAGAGATCTAATGGAAGTGTATGCTTGAATGGAAGATTTAT 558
 QY 248 TTTTACCTAAAG 307
 Db 559 TTTTACCTAAAG 618
 QY 308 AGGATGAGAGATGAG 367
 Db 619 AGGATGAGAGATGAG 678
 QY 368 TGCACTCTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
 Db 679 TGCACTCTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
 QY 428 GACGAAATTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
 Db 739 GACGAAATTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
 QY 488 GGCCTGCTCACTACTCTCATCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 547
 Db 799 GGCCTGCTCACTACTCTCATCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858
 QY 548 TTACCAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 607
 Db 859 TTACCAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 918
 QY 608 AGACAG 667
 Db 919 AGACAG 978
 QY 668 TTACCAATTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 727
 Db 979 TTACCAATTAATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038
 QY 728 AGAGCACTTTAAG 787
 Db 1039 AGAGCACTTTAAG 1098
 QY 788 AATACCAAG 847
 Db 1099 AATACCAAG 1158
 QY 848 GCGCGGCGCTGGGAG 907
 Db 1159 GCGCGGCGCTGGGAG 1218
 QY 908 TTACAGCTTACTGATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
 Db 1219 TTACAGCTTACTGATTAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1278
 QY 968 ATACAG 1027
 Db 1279 ATACAG 1338

Wed May 12 09:50:20 2004

us-10-043-774b-1.rge

Page 10

ORIGIN

DN SGLKPDHIQR LTYKLCHIY YNWPVSASTC

Query Match	94.5%	Score 2200.2;	DB 9;	Length 2579;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 2202;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0.

QY	8	TTGGTGTGAACACACAGGAGAAACTTGAACATGTGTAAAGATCAAAAACAGGTTCTTGAG	67
Db	319	TTGGTGTGAATACAAAGCGAGAACTTACACATGTTAAAGATCAAAAACAGGTTCTTGAG	378
QY	68	GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGCATCCGTCCTCCACTGGGCGCT	127
Db	379	GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCGTGCATCCGTCCTCCACTGGGCGCT	438
QY	128	TATATCAGTATCACATTGACTATTAACCACTGATGGAAGCCAGAAAGCTCCGTTCAAGCTC	187
Db	439	TATATCAGTATCACATTGACTATTAACCACTGATGGAAGCCAGAAAGCTCCGTTCAAGCTC	498
QY	188	TTCTTTTTCACACGAGAGATCTAATTGGAAGTGTCAATGCTTTTGAATGAACATATTAT	247
Db	499	TTCTTTTTCACACGAGAGATCTAATTGGAAGTGTCAATGCTTTTGAATGAACATATTAT	558
QY	248	TTTTTACCTTAAAGAATAACGAAAGGTACTGAAGTTTATGTAAGACCCGGATATGAG	307
Db	559	TTTTTACCTTAAAGAATAACGAAAGGTACTGAAGTTTATGTAAGACCCGGATATGAG	618
QY	308	AGATGTGAGATTAACATCACTTTAACTAATGAACCTTCCACTACATGCACAACTGTT	367
Db	619	AGATGTGAGATTAACATCACTTTAACTAATGAACCTTCCACTACATGCACAACTGTT	678
QY	368	TGCATGCTATAATATTTTTCAGAGAGCTTTTGAATAATCATGATTTTGCAAAATTG	427
Db	679	TGCATGCTATAATATTTTTCAGAGAGCTTTTGAATAATCATGATTTTGCAAAATTG	728
QY	428	GACGAATTTATTAACCCAAATGACCCATTTGATTTCCAAAGTCAAGGTTGGTGAATT	487
Db	739	GACGAATTTATTAACCCAAATGACCCATTTGATTTCCAAAGTCAAGGTTGGTGAATT	798
QY	488	GAGCGGCGTTACATCTCCATCCTTCAGTATGAAGAAACAGCATATCTCTGCAGTACG	547
Db	799	GAGCGGCGTTACATCTCCATCCTTCAGTATGAAGAAACAGCATATCTCTGCAGTACG	858
QY	548	TTAGCATTAAGTCTTCGAAAGTGAAGCTGTTTGGATTTCATGTTCACTTTATATATC	607
Db	859	TTAGCATTAAGTCTTCGAAAGTGAAGCTGTTTGGATTTCATGTTCACTTTATATATC	918
QY	608	AGACGAAGAAATATAATTTCAAGAAACAAGTTCCAAAGACTAATAGTATTAGTTGTC	667
Db	919	AGACGAAGAAATATAATTTCAAGAAACAAGTTCCAAAGACTAATAGTATTAGTTGTC	978
QY	668	TTACCAAGTATACATATAGACATACAGAGTGGATATTGACTGGGACCAAGATCCCA	727
Db	979	TTACCAAGTATACATATAGACATACAGAGTGGATATTGACTGGGACCAAGATCCCA	1038
QY	728	AAGACACTTTAAGAAAGCCGAGCGCTCTGAAGTACGCTTTTGAATAACTACAGGAAGC	787
Db	1039	AAGACACTTTTAAAGAAAGCCGAGCGCTCTGAAGTACGCTTTTGAATAACTACAGGAAGC	1098
QY	788	AATACAAACCAAGAGATCACCGACTTGAAGCAGCGTCTTGGTCAAGCCAGCCAAAGAA	847
Db	1099	AATACAAACCAAGAGATCACCGACTTGAAGCAGCGTCTTGGTCAAGCCAGCCAAAGAA	1158
QY	848	GAGCGGGGCGCTGGGGGGACATGCGCAAGGCGCTCGCATGCTCATCTCTGCTGATAC	907
Db	1159	GAGCGGGGCGCTGGGGGGACATGCGCAAGGCGCTCGCATGCTCATCTCTGCTGATAC	1218
QY	908	TTTACAGTCTAATCTGATAAATAGCGTAATGATTTTAAGTGAAGAAACATTAAGCGGCTC	967
Db	1219	TTTACAGTCTAATCTGATAAATAGCGTAATGATTTTAAGTGAAGAAACATTAAGCGGCTC	1278
QY	968	ATACAGACTAACTCCAGAGCAAGGCAAGCGTAAGTGGAGCACTCATGATTAATTC	1027

Db	1279	ATACAGACTACTACCCAGACGAAAGGACGGCTGAGAGTGGGAGACACTGATTGATTACATTC	1333
Qy	1028	ATAAAAAGATATGTTCAAAAGGAGCTTCGACCTGGGCTTTGAGCTTTGATTCCACT	1087
Db	1339	ATAAAAAGATATGTTCAAAAGGAGGAGCTTCGAGCTGGGCTTTGAGCTTTGATTCCACT	1398
Qy	1088	TACGTGCTCTCGAGGAAGATTTTGCAACAGAAAAGATTCCCAAGGTGGAATCAT	1147
Db	1399	TACGTGCTCTCGAGGAAGATTTTGCAACAGAAAAGATTCCCAAGGTGGAATCAT	1458
Qy	1148	TTGATTCAATCCAAATTTGACAGATTGCTCCAAAGAACAGAGGTGCACATTATTA	1207
Db	1459	TTGATTCAATCCAAATTTGACAGATTGCTCCAAAGAAAAGAGGTGCACATTATTA	1518
Qy	1208	GTGTTAAGCCACTAGATPACCTGGCTGTGATCTATAGCCGAAAGAAATTAGAACACCA	1267
Db	1519	GTGTTAAGCCACTAGATPACCTGGCTGTGATCTATAGCCGAAAGAAATTAGAACACCA	1578
Qy	1268	ATTCATTGATCAAAATCTATTTTAAGTTACACAGCCATGGCGATSCAAATGAAAAG	1327
Db	1579	ATTCATTGATCAAAATCTATTTTAAGTTACACAGCCATGGCGATSCAAATGAAAAG	1638
Qy	1328	CAATATGATTTGAAGTGAATGACAGAACTGAAGCTTCTPAAGAGCTCTPACAGAAAAG	1387
Db	1639	CAATATGATTTGAAGTGAATGACAGAACTGAAGCTTCTPAAGAGCTCTPACAGAAAAG	1698
Qy	1388	TCACAGCAGACCCAGATAGTGTGCTGCTGTGTCAAGTAATCGNAGGACAAATAG	1447
Db	1699	TCACAGCAGACCCAGATAGTGTGCTGCTGTGTCAAGTAATCGNAGGACAAATAG	1758
Qy	1448	ATGCTATTAAAAATATCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGGCCC	1507
Db	1759	ATGCTATTAAAAATATCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGGCCC	1818
Qy	1508	GAACTTTAGGCAAAAGCAAACTGTCAAGGCATTTGCTACAAAGATTGCCCTACAGATG	1567
Db	1819	GAACTTTAGGCAAAAGCAAACTGTCAAGGCATTTGCTACAAAGATTGCCCTACAGATG	1878
Qy	1568	ACTGCAGATGGGAGAGAGCTCTGGAGGCTGGACATCCCTCTGAAGCTCGTGAATGATCG	1627
Db	1879	ACTGCAGATGGGAGAGAGCTCTGGAGGCTGGACATCCCTCTGAAGCTCGTGAATGATCG	1938
Qy	1628	TTGGGATCGATTTGTTACATGACATGACAGCTGGGCGAGAGCTCATCGCAGATTGTGTG	1687
Db	1939	TTGGGATCGATTTGTTACATGACATGACAGCTGGGCGAGAGCTCATCGCAGATTGTGTG	1998
Qy	1688	CCAGCATCAATAGAGGATGACCCGCTGATTCTACGCTGCATATTCAAGATTAGAGAC	1747
Db	1999	CCAGCATCAATAGAGGATGACCCGCTGATTCTACGCTGCATATTCAAGATTAGAGAC	2058
Qy	1748	AGGAGCTGTTAATGGGCTCAAAAGCTGCTCGCAAGCGGCTCTGAGGCTTGGAAATAGCT	1807
Db	2059	AGGAGCTGTTAATGGGCTCAAAAGCTGCTCGCAAGCGGCTCTGAGGCTTGGAAATAGCT	2117
Qy	1808	GCAATAGTACATGCCCAGCCGAGATCATCGTGTACCGAGTGGCTPAGAGACCGGCAAGC	1867
Db	2119	GCAATAGTACATGCCCAGCCGAGATCATCGTGTACCGAGTGGCTPAGAGACCGGCAAGC	2178
Qy	1868	TGAAAACACTGGTGAACCTACGAAGTGCCACATTTTGGATTGTCTTAAATTCATTGGTA	1927
Db	2179	TGAAAACACTGGTGAACCTACGAAGTGCCACATTTTGGATTGTCTTAAATTCATTGGTA	2238
Qy	1928	GAGGTACCAACCTTAGACTAACGCTAATGTGTGTAGAAAAGAGTGAACACCAAGATTTT	1987
Db	2239	GAGGTACCAACCTTAGACTAACGCTAATGTGTGTAGAAAAGAGTGAACACCAAGATTTT	2298
Qy	1988	TTGCTCAGCTGTGAGAGAACTTCAAAATCCACTTCCTGGAAACAGTTATTGATGTAGAGG	2047
Db	2299	TTGCTCAGCTGTGAGAGAACTTCAAAATCCACTTCCTGGAAACAGTTATTGATGTAGAGG	2358
Qy	2048	TTACACAGACAGAAATGATGACTTTTATTCGTGAGCCAGGCTGTGAGAAAGTGTAGTG	2107
Db	2359	TTACACAGACAGAAATGATGACTTTTATTCGTGAGCCAGGCTGTGAGAAAGTGTAGTG	2418

Wed May 12 09:50:20 2004

us-10-043-774b-1.rge

Page 11

QY	2108	TTTCTCCACACATTAACAATCATCTCATCTCTATGACAAACAGGCGCTGAAGCCACACACATAC	2167
Db	2419	TTTCTCCACACATTAACAATCATCTCATCTCTATGACAAACAGGCGCTGAAGCCACACACATAC	2478
QY	2168	AGGCGTTGACCTACAAAGCTGTGCCAATCTCTTTTCAATGTGGCGAG	2212
Db	2479	AGGCGTTGACCTACAAAGCTGTGCCAATCTCTTTTCAATGTGGCGAG	2523

RESULT	7			
AB032604				
LOCUS	AB032604			
DEFINITION	Mus musculus mRNA for MMTV, complete cds.	3910 bp	linear	ROD 04-OCT-2001
ACCESSION	AB032604			
VERSION	AB032604.1	GI:7416110		
KEYWORDS	MMTV (piv1).			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			

REFERENCE
AUTHORS
1 Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A.,
Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y. and Nakano, T.
TITLE
Two mouse p1v1-related genes: m1v1 and m1l
JOURNAL
Mech. Dev. 108 (1-2), 121-133 (2001)
MEDLINE
21463379
PubMed
11578866
2 (bases 1 to 3910)
REFERENCE
AUTHORS
Miyagawa, S.K. and Nakano, T.
TITLE
Direct Submission
JOURNAL
Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for

FEATURES	source	Location/Qualifiers
		1..3910
		/organism="Mus musculus"
		/mol_type="mRNA"
		/db_xref="taxon:10090"
		/map="5F distal-G3 proximal"
		/tissue_lib="testis"
gene		1..3910
		/gene="miw1"
CDS		156..2744

ORIGIN				
Query Match	72.4%	Score 1684.6;	DB 10;	Length 3910;
Best Local Similarity	82.8%	Pred. No. 0;		
Matches 1924;	Conservative	0;	Mismatches 399;	Indels 0;
			Gaps	0;

OY	6	CTTTGGTGTGAACAACAGGACAGAACTTAAAGACATGTGTAAGAAATCAAAAAAGGCTTCTC	65
Db	422	CTTTGGTGTGAACAACAGACAGAAACCTTGAACATGTCAAAAGATCAAAAGACAGGCTCTCC	481
OY	6	AGGCATTATATGTAAAGTTAAAGCATACATTTCCGGCTCAATCCCGTCCCAAGTGGCG	125
Db	482	TGGCATATTGTGAAGCTGAGACCAACCACTTCGCGCTACCTCGGCCCAAGTGGCG	541
OY	126	CTTATATCAGTATCACTTACTATATAACCACTGATGGAAGCAGAAAGATCCGTTCAAG	185
Db	542	CTGTATACAGTACCAATGACATGACTATATCCCTGTATGGAAGCCGGAAGGCTTGCGCTCGC	601
OY	186	TCTTCTTTTTCACACGAAATCTAATTTGGAAGGTGCATGCTTTTGTATGGAAGGATAT	245
Db	602	ACTGCTCTTCACAGATMAAGCTCATTTGGAAGGTGCATGCTTTTGCATGGACATATTT	661
OY	246	ATTTTTCCTAAAAAGCTACAGCAAAAAGTTTACGTAAGTTTATAGTAAGACCCGGAAATG	305
Db	662	GTTTTTCCTTAAGAGCTACAGCAAAAGTACAGAAAGTATTCAGTCACTCGGAATGG	721
OY	306	AGAGAGATGTAGATACGATCACTTTAACAATGAACTTCCACCTTATATCAACCAATTG	365
Db	722	GGAACAGGTAGAGATCCCATCACTCCGACCAACAGACTCCGCCCACTGCGCCACCTG	781
OY	366	TTTGACGTTCTATATATTTTTCAGAGAGGCTTTTGAATATGATGATTTTGCACAAAT	425
Db	782	CTTGCAATTTCTATATACATCATCTTCAAGAGGCTTTGAATATCATGATTTTGCACAAAT	841
OY	426	TGACGAAATTTATATACCCAAATGACCCATTTGATATTTCCAAGTCAACAGTTTGTGAT	485
Db	842	TGACGGAATTTATCAATCCAAAGTGCACGATTTGATTTTCCAAACCAACAGTTTGTGAT	901
OY	486	TTGACCTGGCTTCACTACTTCCACTCTTCAGTATGAAACAGACATCATGCTTGCATGTA	545
Db	902	CTGCGCGGCTTCAACACTCCATCCCTTCAGTATGGAACAATCATGCTTGCACAGA	961
OY	546	CGTATGCCATTAAGTCTTTGGAAGTGAAGCTGTTTGGATTTCAATGTTCACTTTATCA	605
Db	962	CGTACGCCAACAGGTGCTCCGACGACGAGCTGTCCTTAAGTTCATGTTCAATCTATACCA	1021
OY	606	TCAGACGAAACATATAATTTCAAGAACAGTTTCCAAAGAACTAATAGTTTATGTTGT	665
Db	1022	GCAGACGAGAGACAAAGTTCCAGAGCAAGTGTGAAGAGTGTATATAGGCGTCATGCT	1081
OY	666	TCTTACCAAGTATTAACAATTAAGACATACAGATGAGATGATTTGACTGGGACCCGAATCC	725
Db	1082	TCTACCAAGTATTAACAAGACCTACCGGATGAGATGACATTTGACTGGGACCCGAATCC	1141
OY	726	CAAGAGACCTTTTAAGAAACCCACGCGCTGGAAGTCAAGCTTCTTAAGATCTACAGAA	785
Db	1142	AAAGAGACCTTTCAAGAAAGGCGATGCGTGGAGTCAAGCTTCTTGAAGTATCTACAGAA	1201
OY	786	GCAATACACCAAGAGATCAACGACTTGAGACAGCTGTCTTGTCAAGCCACGCCAAG	845
Db	1202	GCAATACACCAAGAGATCAACGACCTTGAGACAGCCGCGTCTGTGACCAACCCAAAGCG	1261
OY	846	AAGCGGGGGCCCTGCGGGGACATGCGCAAGGCGCTGCATGCTCTTCTGAGCTTGTCTA	905
Db	1262	GAGAGAGAGCCCGCGCGGACCTGCGCTGCGCCAGATATGCTATCCCTGAACTGTCTA	1321
OY	906	TCTTACAGGTCTAATCTATATAATGCGTATGATTTTAAAGTATGAAATAAGCTTAAGCGGT	965
Db	1322	TCTTACAGGCTGTACTATATAATGCGCAATGATTTTAAAGTATGAAAGACTTGGCAAT	1381
OY	966	TCATACAAAGCTAACTCAAGACAAAGGACGCTGAAGTGGAGCACTCATTTGATTCAT	1025
Db	1382	GCAACGCGGTGACCCCTGAGCAGCGGACGCGGAAGTGGGCGGCTCATGCACTCAT	1441
OY	1026	TCATAAAAAGATATATGTTCAAAAGGAGCTTTCGAGATCGGAGGTTTGAAGCTTTGATCCAA	1085
Db	1442	CGAACAAGATGACATATGTGACGAAGAGCTTCGAGATCGGAGGCTTGAAGCTTTCGATCAA	1501

QY	1086	CTTACGTGCTTCTCGAGGAAGATTTTGGCAACAGAAAAGATTCACCAAGGTGGAAAAC	1145
Db	1502	CTTGCTGCTCTCTCTGGAAGATCTTCAATCTGAGAGATCCACGAGCGGAAGAC	1561
QY	1146	ATTGTATTCAATCCAAATTTGCAATGGTCCMAAGAAACAAGGTGCACATTAAT	1205
Db	1562	GTTGTATTCAACCCCAATTTTGCAAGCTGGTCCAAAGAAACAAGGCCGCGCTGAT	1621
QY	1206	TAGGTTAACCACTAGATACTGGCGTGTGATCTAAGCGGAAGAAATTAAGACAG	1265
Db	1622	CAGGTGAACCCATGGATTAACGTGGCTGTGATCTAACCAGGAATTTAAGACAGC	1681
QY	1266	CAATTCATTATACMAAATCTATTTAAAGTTACACAGCCATGGGCATGCAATGAGAA	1325
Db	1682	CAATCTCATGTATACGAACCTGTTCAAAAGTACTCCAGCCATGGGCATCAGATGAAAA	1741
QY	1326	AGCAATATAGATTGAAGTGAATGACAGAACTGAAGCTACTTAAAGTCTTTAAGCAAAA	1385
Db	1742	GGCAATCATGATCGAGGTGATGACAAACAAGGCTTAATCTGAAGCTTTGACAGAAA	1801
QY	1386	GGTCAAGCAAGACCCAGATAGTTGTCTGTCTTGTCAAGTATCGAAGACAATA	1445
Db	1802	GGTAACTCGACACCTCAGATAGTTGTCTGTCTTGTCAAGTATCGAAGACAATA	1861
QY	1446	CGATGCTATTTAAAAATACCTGTGTACAGATTGCTCTTACCCCAAGTCAGTGTGTGGC	1505
Db	1862	TGATGCCATCAAGAAAGTACTGTGTACAGATGCCCAACCCCAAGTCAGTGTGTGGGC	1921
QY	1506	CCGAACCTTAGGCAAAACAGCAAACTGTCAATGGCCATTTGTCACAAAGTTGCCCAAGAT	1565
Db	1922	CCGAACCTTAGGCAAAACAGCAAAAGTATGGCCATTTGTCACAAAGTTGCCCTCGACAT	1981
QY	1566	GAACTGCAGAATGGAGAGAGAGCTCTGAGGGGTGACATCCCGCTGAAGCTCGTATGAT	1625
Db	1982	GAACTGCAGAATGGAGAGGCGAGCTCTGAGGGGTGACATGCGCTGAAGCTCGTATGAT	2041
QY	1626	CGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGAGGTGCATTCGAGAGATTGT	1685
Db	2042	CGTGGGCTACACACTGTTACCATGACACCAAGCTGGGCGAGGTGCATTCGAGAGATTGT	2101
QY	1686	TGCCAGCATCAATGAAGGGATGACCCCGCTGCTTCAAGCTGCATATTTCCAGATPAGG	1745
Db	2102	CGCCAGCATCAATGAAGGGATGACCCCGCTGCTTCCCGCTGCGCTTTACGACCGGG	2161
QY	1746	ACAGAGCTGTGATGATGGGCTCAAAAGTCTGCTCGACAGCGCTCTGAGGGCTTGAATAG	1805
Db	2162	GCAAGAGCTGTGATGATGGTCTCAAGGTGTGCTTGCAGCTCTCTGAGGGCTTGAATAG	2221
QY	1806	CTGCAATAGATACAGGCCACGCGGATCATGCTGTACCGCATGGCGATGAGAGACGGCCA	1865
Db	2222	CTGCAATAGATACAGGCCAGCCGCTGTATCTGTACCGGAACGCTGTGGGGAGCGGGCA	2281
QY	1866	GCTGAAAACACTGTGAACTACGAAGTGCACAGTTTGTGATGTCTAAATCATTTGG	1925
Db	2282	GCTGAAAGCCCTGGTCAATTAATGAGGTGCCACAGTCTCTAATGTGCTCAAGTCAGTGG	2341
QY	1926	TAGAGGTTACAACCTTAGACTAAAGGTATGTGGTGAAGAAAAGAGTGAACACAAGATT	1985
Db	2342	GAGAGGTTACAACCCAGACTGACTGTATGCTGGTGAAGAAAGGTGCATATGCCAGTT	2401
QY	1986	TTTTGCTCAGTCTGAGAGAGACTTCAGAAATCGACTTCTTGAAACGATTATGATGAGA	2045
Db	2402	TTTTGCTCAGTCTGGGGAGAGACTTCAGAAACCTCTTCCAGGGAGACGTCACTGATGTGA	2461
QY	2046	GGTTAACGACCAAGATGTATGACTTTTATATCGTGAAGCAAGGTGTGAAGAAGGTAG	2105
Db	2462	AGTCAACCAAGACGAGGTGTATGACTTTTATATCGTGAAGCAAGGTGTGAAGAAGGTAG	2521
QY	2106	TGTTTCTCCCAACAATTACATGTGATCTATGACAAACGCGCTGAACCAACACAT	2165
Db	2522	TGTTGTCCCAACAACACTACATGTATCTATGACAGAGGTGCTGAAGCCGACACAT	2581
QY	2166	ACAGGCTTGACCTACAGCTGTGCCATCTATTACATGGCCAGAGTGCATTCCGT	2225

Db	Accession	Score	Length	Gap	Indel	Mismatch	Match	Similarity	Conservative
Db	2582	CCAGCGGCTACATACAGTCTTGCCACGCTGACTAATTAATGGCCCTGAGTATCCGAGT	2641						
Qy	2226	TCCTGCTCTTGGCAGTACGCCCCAAGCTGGCTTTTCTTTGGCCAGATATTACAG	2285						
Db	2642	CCCTGACCTTGGCAGATATGACACAGAGCTGGCTTCCTGTTGGGCGAAGCATCCAG	2701						
Qy	2286	AGAGCCCAATCTGTCACCTGTCACACCGCCCTTACTACTACTCTTAA	2328						
Db	2702	AGAGCCCAACCTCTCTCTGTCACACCGCTCTTACTACTCTTAA	2744						
RESULT 8	AF438405	4064 bp	linear	ROD 13-NOV-2001					
LOCUS	AF438405	Mus musculus M171 mRNA, complete cds.							
DEFINITION	AF438405.1	GI:16905060							
ACCESSION	AF438405.1	GI:16905060							
VERSION	AF438405.1	GI:16905060							
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
CDS									
ORIGIN									
Query Match	72.4%	Score 1684.6	DB 10	Length 4064					
Best Local Similarity	82.8%	Pred. No. 0							
Matches 1924	Conservative	0	Mismatches 399	Indels 0	Gaps 0				
Qy	6	CTTTGGTGTGAA CACAAGCAGAACTTGAACCATTTAAAGATCAAAAACAGGTTCTTC	65						
Db	457	CCTTGATGTGAA CACAAGCAGAACTTGAACCATTTAAAGATCAAAAACAGGTTCTTC	516						
Qy	66	AGCAATTAAGTAAGTTAAGACATCAATTCGCGTGAACATCCCTCCCAAGTGGAC	125						

Db	1459	TTTGAGTAAACCATGACAGACAGACTGGCTCTAAAGAGATGAGAGGTCTCCCTCGATT	1518
Qy	1207	AGGTGTAAGCCATGTAATTAATGGGTGTGATCTTAACGGAGAAATTAATGAGCAACC	1266
Db	1519	AGTGCAGTGTCTTGAGCAACTGGGTCAATGTTCTACACGGCGAGAAATGCAAGATGTCC	1578
Qy	1267	AATTCATTGATACAAATCTATTTTAAATGTAACACGGCATGCGGATCAATGAGAAA	1326
Db	1579	CAATCTCGTTCAAGACTCTCAACAAATGTCTGAGCCATGGGGATCCGATGACAGG	1638
Qy	1327	GCAATTAATGATTTGAAGTGAATGACAGAACTGAAGCTCTTAAGACTTTACAGAAAG	1386
Db	1639	GCTGCAGATGATGATGAAGATGTCAGAAATCTCTTAAAGCTTTGCACAAAAAC	1698
Qy	1387	GTCACAGACAGACCCAGATAGTTGTCTGTCTGAAGTAATCGAAGAGACAAATAC	1446
Db	1699	GTAGCACTGAAATCTCAATAGTGTGTGATCTTGCTTACCAATGCAAAAAGCAAGTAC	1758
Qy	1447	GATGCTATTAATAAAATACCTGTGTGACAGATTGGCCCTAACCCCAAGTCAGTGTGTGGCC	1506
Db	1759	GACTGTGTAAAGAAAGTACTCTGTGTGTGACTGGCCCTTACCTCAAGTCAAGTGTGTGTC	1818
Qy	1507	CGAATCTTAAGCAAAACACAAATCTGTCAATGCGATTTGTACAAATATGGCCCTACAGATG	1566
Db	1819	CGCACATACAGAAACTTCAAGCACTTAATGATGTGGCCACCAAAATTTGCTTGACAGATG	1878
Qy	1567	AACTGCAGATGAGGAGAGAGCTCTGAGAGGTGACATCCCTGMACTGTGATGATC	1626
Db	1879	AACTGCAGAAATGGGAGAGAGAGTGTGAGTGTGAATTCACATCGGCAGCTCATGATT	1938
Qy	1627	GTTGGCATTCGATTTTACATGACATGACAGTGTGGCGAGAGTCATATGCAAGATTTGTT	1686
Db	1939	GTAAGTAATGACTGTCTACCATGATATCCGCTGTGTGAAAAGGTCATGTGGCATGTGTG	1998
Qy	1687	GCCAGCATCATGAAAGGAGTATCCCGCTGGTTCACCGCTGATATTTACAGATAGAGGA	1746
Db	1999	GCCAGTCTTAATCAGGCGATGTCAAGATGTGTTCTCCAGTGTGTCTTGCAAAACCGTGA	2058
Qy	1747	CAGAGCTGTGATGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGAAATAGC	1806
Db	2059	CAGAGATCATGCAATGCACTGAAAGGGGTCAATTGCAAGGGTGTCTGMAAGCCATCTGAAA	2118
Qy	1807	TGCAATGATGATACATCCCAAGCCGATCATCTCGTATCCGCGAATGGGTATGAGAGCGCAG	1866
Db	2119	TACAACTAATCTTGCCCTCGGCAATATGTTTACAGAGTGAAGTTGAGATAGCAATG	2178
Qy	1867	CTGAAAACACTGATGAATCAAGAGTGGCAAGTTTGTGATTTGATCTTAAATCCATTGCT	1926
Db	2179	CTGCAGACGCTTGCGACTATATATATCTCTCAGATCATGCACTATCAAGCAATGGGG	2238
Qy	1927	AGAGGTTCAACCTTGAATCTAACCGGTAATTTGTGTGAAGAAAGATGAACACAGATTT	1986
Db	2239	CAAGATTATGAGCCCAAGCTCTCACTATATGTGTGAAAAAGCGATCAAGCTCCCGATT	2298
Qy	1987	TTTGTGCTAGTCTGAGAGAAAGATTGAGATCCATCTCTGAAACAGTATTTGATGATAGAG	2046
Db	2299	TTTGTGTGGATTGATGAGCAAAATTTGCCAATCCCTCTCGGAGACAGTCAATGACCTGAA	2356
Qy	2047	GTTACACAGACCAAAATGATATGACTTTTATCTGTAGCCAGGCTGTGAGAAATGTATGT	2106
Db	2359	GTCATCTGTCCAGATGTGATATTTCTTCATTTGTAACCAAGGCTGTGCGCTTGGGGTGT	2418
Qy	2107	GTTTCTCCCAACATTACAAATGTCACTATATGACACAGGGGCTGAAGCCAGACCACTA	2166
Db	2419	GTGCACTTAATCTCACTCAATATGTGTGTTGACACAGCGGCTTTAAACAGATCAATG	2478
Qy	2167	CAGCGCTTGACCTTACAGACTGTGCAACATTTTACAACTGGCCAGGTGTCTTCGTGT	2226
Db	2479	CAGAGACTCACTTACAACTTTGCAACATGTACTCAACTGGCAAGGATTTGTGAGATG	2538
Qy	2227	CTGTGCTCTTGCAAGATAGCCCAAGAGCGGTCTTTCTTTGTGGCAAGTATTCACAGA	2286
Db	2539	CAAGTCTATGCAATATAGCCCAAAATTTGGCTTCTGTGTGTGGCCAGACATTCACAG	2598

Db 2991 GGTGGCTTACTCGTGGGGGAGTCATTCACACAGGAACCAATGCTTCTGTCACTCG 3050
 QY 2313 CCTTACTACTCTTAA 2328
 Db 3051 TCTCTTACTCTTGA 3066

RESULT 11
 BC031060 3006 bp mRNA linear PRI 06-OCT-2003
 LOCUS Homo sapiens piwi-like 2 (Drosophila), mRNA (cDNA clone MGC:35296
 IMAGE:5271278), complete cds.
 ACCESSION BC031060
 VERSION BC031060.1 GI:21410560
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3006)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Nak, S.I., Wang, J., Hsieh, F.,
 Diachenko, L., Marisina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stachleiron, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheet, T.E., Brownstein, M.J., Uedl, T.B., Toshitsuki, S.,
 Carinoni, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.C., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Heltun, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bonfield, W.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smalls, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL 22388257
 MEDLINE 12477932
 PUBMED 2 (bases 1 to 3006)
 REFERENCE Strausberg, R.
 TITLE Direct Submission
 AUTHORS Submitted (03-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshitsuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: http://www-shgc.stanford.edu
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

FEATURES
 source
 1..3006
 /organism="Homo sapiens"
 /mol_type="mRNA"

misc_feature
 817..1224
 /note="PZ; Region. PZ domain. This domain is named PZ
 after the proteins Pwi Argonaut and Zwill. This domain
 is found in two families of proteins that are involved in
 post-transcriptional gene silencing. These are the Pwi
 family and the Dicer family, that includes the Carpel
 factory protein. The function of the domains is unknown
 but has been suggested to mediate complex formation
 between proteins of the Pwi and Dicer families by
 hetero-dimerisation"
 /db_xref="CDD:pfam02170"
 1639..2517
 /note="Pwi; Region. Pwi domain. This domain is found in
 the protein Pwi and its relatives. The function of this
 domain is unknown"
 /db_xref="CDD:pfam02171"

misc_feature
 1639..2517
 /note="Pwi; Region. Pwi domain. This domain is found in
 the protein Pwi and its relatives. The function of this
 domain is unknown"
 /db_xref="CDD:pfam02171"

ORIGIN
 Query Match 29.2%; Score 680; DB 9; Length 3006;
 Best Local Similarity 56.7%; Pred. No. 4.5e-172;
 Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

QY 1 ATGATCTTGGTGGTGAACACAGGACGAACTTGAACCTTAAAGATCAAAACAGGT 60
 Db 244 ATGATTTGGATGATCTGTACACAGAAATTTGACATGTGAATTTTAAACAGGT 303
 QY 61 TCTTACGACATTTAGTAAGTTAAGACATTAACATTTCCGGCTGACATCCGTCGAC 120
 Db 304 TCCAGTGAATCTGCTGAAATCGTTACAAACCTTTTAACTGATTTTCCCAAGAC 363
 QY 121 TGGGCTTATATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 Db 364 TGGGAGCTATACAGTACATGATGATGATGATGATGATGATGATGATGATGATGAT 423
 QY 181 TGAGCTCTCTTTTCAACAGAAAGATCTAATGAAAGTGTGATGCTTTTGAAGAC 240
 Db 424 ATTGCTTACTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
 QY 241 ATATATTTTATACCTAAAGACTACAGCAAAAGGTTAAGATTTTATGATGACCCGG 300
 Db 484 ATCTTTTCTGTCACAAAGAGTGAAGAAAGGTGACAGAGTTGTGATGATGAAATCA 543

QY 301 AATGAGAGAGTGAAGATTAACATCACTTTAAACAATGAACTTCCACTACATACCA 360
DB 544 AGAGGTGAGACTTAAGATGACTTACACCTGAAAGAGAGGTGCAATCAAGTTCTCC 603
QY 361 ACTGTTGAGATTTCTAATAATTAATTTTCAAGAGCTTTGAAATCAATGATTTGCA 420
DB 604 GTGTCATCCAGGCTTCAATATCATCTTCAGAAAGATCTCTAAAAGTTGTCATGAC 663
QY 421 CAAATGAGAGAAATTTATTAACCCCAATGATGATTAATTCAGATGACAGGTTG 480
DB 664 CAAATGAGAGAACTTCTATTAATCTTCAAGAGCAATGAAATTTCCACAGCAATTA 723
QY 481 GTGATTTGGCTGAGCTTCACTACTTCCATCTTCAATGATGAAACAGCATCATGCTG 540
DB 724 TCCCTTTGGCTGAGGTTGCAATTTCTGTCATATTTGAAAGAGAGCTCTGTTAAT 783
QY 541 ACTGACGTTAGCCATAAGCTCTGAAAGTGAAGCTGTTGATTTCAATGTTCACTTT 600
DB 784 GCTATGATGATTAAGATGATCTCCGAAATGAGAGCTTCTGAAATTCATGACTGCTC 843
QY 601 TATCATCAGACAGAGAACTAAATTTCAAGAACTTTCCAAAGACTTAATAGTTTA 660
DB 844 TGTCAAAGAACTGGCTTGTCTGTTTCAACCAAGCTGTGAGAAAGCTTAATAGAGCTC 903
QY 661 GTTGTCTTCAAGATTAACATTAAGACATTAAGAGATGATTAATGACTGAGACAG 720
DB 904 ATTGTCTTCAAGATTAACATTAAGACATTAAGAGATGATTAATGACTGAGACAG 780
QY 721 AATCCCAAGAGACCTTTAAGAAAGCCGAGCTCTGAAAGTCAAGCTTCTTAAGATTA 780
DB 964 AAGCCCAACACACCTTTCAAGAGAGGAGTGGCAACGAGATCACTATGATGATTAAC 1023
QY 781 AGAAGACATTAACCAAGAGATCAACGACTTGAAGAGAGCTGCTGTCAGAGAGCC 840
DB 1024 AAGAGAGAGTATGATTAATTAATGATGAGACCTGATTAAGAGAGCTGTTGATG 1083
QY 841 AAG 900
DB 1084 AAG--AAG 1140
QY 901 TGTATCTTCAAGAGCTTCAAGATTAATGATTAATGATTAATGATTAATGATTA 960
DB 1141 TGTCTTCAAG 1200
QY 961 GCGCTTCAACAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 1201 GGTGAAAG 1260
QY 1021 TACATTCATTAAG 1080
DB 1261 AAGATTCAG 1320
QY 1081 TCCAACTTACTGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1321 AAGCCAGATTA--TCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
QY 1141 AAG 1200
DB 1378 CACATTAAT--GTCAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434
QY 1201 TTAATTAATGATTAAG 1260
DB 1435 ATTTTAATGAG 1494
QY 1261 GAG 1320
DB 1495 GTTGGAG 1554
QY 1321 AGAAG 1380
DB 1555 GAGTACCCCAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
QY 1381 CAAAG 1440

DB 1615 CAAATATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1674
QY 1441 AATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
DB 1675 TATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1734
QY 1501 GTGAG 1560
DB 1735 CTGCTGAG 1794
QY 1561 CAGATGAAG 1620
DB 1795 CAGATGAAG 1854
QY 1621 ATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
DB 1855 ATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1914
QY 1681 TTTGTTGAG 1740
DB 1915 TCCGTTGAG 1974
QY 1741 AG 1800
DB 1975 AATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2034
QY 1801 AATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
DB 2035 TACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2094
QY 1861 GAG 1920
DB 2095 GGTGAG 2154
QY 1921 ATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
DB 2155 TCCAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2214
QY 1981 AGATTTTGTCTGAG 2040
DB 2215 CAGATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2274
QY 2041 GTGAG 2100
DB 2275 TCGAG 2334
QY 2101 GGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
DB 2335 GGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2394
QY 2161 CAGATGAG 2220
DB 2395 CAGATGAG 2454
QY 2221 CGTGTCTGAG 2280
DB 2455 AGTGTCCAG 2514
QY 2281 CAG 2328
DB 2515 CATTAAG 2562

RESULT 12
AX748151 3138 bp mRNA 1linear PAT 20-JUN-2003
LOCUS AX748151
DEFINITION Sequence 1676 from Patent EP1306455.
ACCESSION AX748151
VERSION AX748151.1 GI:32132539
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

QY 601 TATCATCAGACAGAAACATTAATTTCAGAAACAGTTTCCAAAGAACTAATAGGTTA 660
 DB 992 TCTCAAGAACTGGCTTGTCTGTTCACCCAGAGCTGTGAGAAACACACTAATAGGGCTC 1051
 QY 661 GTTGTCTTCAAGATTAACAATAAGACATACAGAGGATGATTAATGACCTGGAGCAG 720
 DB 1052 ATTGTCTTCAAGATTAACAATAAGACATTAATGATTAATGATTAATGATTAATGAT 1111
 QY 721 AATCCCAAGACACCTTTAAGAAAGCCGCTGTGAAGTCAAGCTTTTGAATATATAC 780
 DB 1112 AAGCCCAACACACCTTTAAGAAAGCCGCTGTGAAGTCAAGCTTTTGAATATATAC 1171
 QY 781 AAGAAACATTAACAACCAAGATTAACCCGCTTTGAAGACCTGTGTGTGTGACCAAGCC 840
 DB 1172 AAGCAAGCAATGATTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1231
 QY 841 AAGAAAGAGCGGGGCTTGGGGGCACTGCGCAAGGCTGCTGCTGCTGCTGCTGCTGCTG 900
 DB 1232 AAG---AAGAAAGAAATGACAAACAGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1288
 QY 901 TGTATCTTACAGGCTTAATGATTAATGCTGATTAATGATTAATGATTAATGATTAATG 960
 DB 1289 TGTATCTTACAGGCTTAATGATTAATGCTGATTAATGATTAATGATTAATGATTAATG 1348
 QY 961 GCGGTCATTAACAAGCTTAATGCTGAGCAAGGCAAGGCTGAGTGGAGCACTGATTAAT 1020
 DB 1349 GCTGAAAGACAGCTTCACTGCTTCAAGGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTG 1408
 QY 1021 TACATTCATTAACAAGCTTAATGCTGAGCAAGGCTGAGTGGAGCACTGATTAATGAT 1080
 DB 1409 AACATCCAGAGAAATCAATGCTGCTTGAATGAGAACTGAGGAGCACTGATTTTGA 1468
 QY 1081 TCCAACTTACTGCTTCTGAGAGAAATTTTGCAAAACAGAAATGATCAACAAGTGA 1140
 DB 1469 AGCAATA---TCTGAGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1525
 QY 1141 AAAACATTTGATTAACAATTTTGAGATTGGTCCAAAGAAACAGAGGTCACCA 1200
 DB 1526 CACATTT---GTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1582
 QY 1201 TTAATTAAGTTAAGCACTGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 DB 1583 ATTTTAAATGCAAGCTTGAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1642
 QY 1261 GCAAGCAATTTGATTAACAATTTAATTAAGTACCAAGGCTGAGGCAATG 1320
 DB 1643 GTTCCGAGAGCTTCTGAGCTGCTGAGAAAGATTTGCTGCTGCTGCTGCTGCTGCTG 1702
 QY 1321 AGAAAGCAATAATGATTAAGTGAAGTGAAGAACTGAAGCTTACTTAAGCTTTACG 1380
 DB 1703 GACTACCCCAATTAATTAAGTGAAGAAATCCAGCTGCTGCTGCTGCTGCTGCTGCTG 1762
 QY 1381 CAAAAGTCAAGCAACCCGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
 DB 1763 CAATATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1822
 QY 1441 AATTAAGATGCTTAAATAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 DB 1823 TATTAATGATTTCAATTAATAATTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1882
 QY 1501 GTGCGCCGAACTTTAGCAACAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 1883 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942
 QY 1561 CAGATGAATGCAAGATGAGAGAGAGCTGCTGAGGCTGAGCACTCCCTGAGCTGCTG 1620
 DB 1943 CAGATGAATGCAAGATGAGAGAGAGCTGCTGAGGCTGAGCACTCCCTGAGCTGCTG 2002
 QY 1621 ATGATGTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
 DB 2003 ATGATGTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2062
 QY 1681 TTGTTGCAAGATGAAGGATGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

DB 2063 TGCCTGCGCAGGTTTAACCCAGAAATCAAGAGTGTGTTTCCGCTGATCTTCAGAGA 2122
 QY 1741 AGAGCAAGAGAGCTGTGATAGTGGCTCAAAAGTGTGCTGCTGCAAGCGCTGAGGGCTTG 1800
 DB 2123 ACATATGATGATGTGAGATGCTGTTGAAGTTTCTGATGCTGAGCACTCAACAAATGG 2182
 QY 1801 AATGCTGCAATGATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 DB 2183 TACAGATCAATATGATGTTTGCAGAGCAGGATTAATGATGCTGCTGCTGCTGCTG 2242
 QY 1861 GGCCAGCTGAAACACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 DB 2243 GCTGAGCTGAAACACTTATGATTAATGATGATGATGATGATGATGATGATGATGATG 2302
 QY 1921 ATTGTGAGGTTTCAACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1980
 DB 2303 TCCGCTCAATATCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2362
 QY 1981 AGATTTTGTGCTGATGCTGAGAGAACTGATGATGATGATGATGATGATGATGATGAT 2040
 DB 2363 CGATTTCTTACCGAATGAACCGCATGTCAGAAACCCCACTTGGCACTGTTTGAT 2422
 QY 2041 GTAGAGTTTACAGACCAAGATGATGATGATGATGATGATGATGATGATGATGATG 2100
 DB 2423 TCGAAGCAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2482
 QY 2101 GGTATGTTTCTCCCAACATTAATGATGATGATGATGATGATGATGATGATGATGAT 2160
 DB 2483 GGAATGTTTCTCCCAACATTAATGATGATGATGATGATGATGATGATGATGATGAT 2542
 QY 2161 CACATCAAGGCTTGAACCTTCAAGCTGCTGCAATCTTATTAACAAGTGTGAT 2220
 DB 2543 CATATGAGAGACTTATTAATGATGATGATGATGATGATGATGATGATGATGATG 2602
 QY 2221 CGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 DB 2603 AGTGTCCAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2662
 QY 2281 CACAGAGGCAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2340
 DB 2663 CATTAAGACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2720

RESULT 14
 AB079366 3158 bp mRNA linear PRI 02-AUG-2003
 LOCUS Homo sapiens PIV1L1 mRNA, complete cds.
 DEFINITION AB079366
 ACCESSION AB079366.1 GI:33411129
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Sasaki, T., Shiohama, A., Minoshima, S., and Shimizu, N.
 Identification of eight members of the Argonaute family in the
 human genome
 JOURNAL Genomics 82, 323-330 (2003)
 2 (bases 1 to 3158)
 AUTHORS Shimizu, N., Minoshima, S., and Sasaki, T.
 DIRECT SUBMISSION
 Submitted (04-FEB-2002) Nobuyoshi Shimizu, Keio University School
 of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
 160-8582, Japan (E-mail: nshimizu@med.keio.ac.jp,
 Tel: 81-3-3351-2370, Fax: 81-3-3351-2370)
 LOCATION/Qualifiers
 1..3158
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="11"

FEATURES

source

```

gene
1..3158
/seq="PIWILL1"
172..2730
/codon_start=1
/protein_id="BAC81341.1"
/db_xref="gi:33411130"
/translation="MSGARVAKIARSPGAEVGRIOASPLPRSVLDSNENASSN
GFLGSRISTNDKYGISGDAGSTFMERGVKNODFMDLICTREKLAHRNKAQSS
GIPVLTNLNLDLPDMOYOHVYIPLASRLIMLVSHSLSKARAFDA
ILFSLKKEKVEVSEISSTGRTIKMTITLRELPSSEVPCIOFENIIPRLIKLKS
MYOIGRNFYSEMEIPIHKLILMPGPAISVSEKELPSAASVXYLVNGLVLEF
MTALCORGLSCTFOCTCEKULIGLIVLTRNNRTYSIDIDMSVKTPTHEOKRQREI
TYVDYIKQIDITVSDNLQFVLVSLKKRDNSEBOLAHILPELCTGLTLOQATSD
FOLMKAVAEKRLSPSGQRLARLVNDIQNTNARELEHMGHPSQSILGRIVP
SEKILMODHIQPVSAADWSIDIRCKILNOSLNTWLILCSDETVASPLNCLR
VAGSGENVDPKIIKVOENPAFVRAIQOYVDPVOLVMOCLIFSNKTYDSIKYLI
SSDCVPSQCVLARTLNKQGMMSIATKIMQTCIKLGEIWAYEIPLKSIMVGVIDV
CKDALSQVNVVGVASVNPRIITRMFSCLIORMTVDLCKFMTGALKNKYKYNH
DLPAIIVYRAGVGGIKTLIEYVDPILSVASSSNTSSLSVIVVRKCPREF
TEBARTVONPPLGTVDBSEATRNEMVDEPLISVACGTSPTVYVYIDNGIKPDH
MORLFLKCHLITMPEIVSVAPFQIAHKLFLVAGSIIKEPSLEIHLNHLFI"

```

ORIGIN

```

Query Match      29.2%; Score 680; DB 9; Length 3158;
Best Local Similarity 56.7%; Pred. No. 4.5e-172;
Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

```

```

QY 1 ATGATCTTTGGTGTAAACAAGCAGAACTTAAACCTGTTAAAGATCAAAAACAGT 60
DB 412 ATGATTTGAGATCTGTACAGAGAAATTTGACATGTGAGAAATGTAAAGCAGT 471
QY 61 TCTTCAGGATATAGTAAGTTAAGCACTAACCAATTCGGCTGACATCCCGTCCAG 120
DB 472 TCCAGTGAATACCTGTGAAACTGGTTACAAACCTTTAACTTGAATTTCCCAAGAC 531
QY 121 TGGGCTTATATCACTATCACTTAACTTAAACCACTGATGAAACCAAGAACTCCGT 180
DB 532 TGGCAGCTATACCAATGATGTGACATATATCCAGATTTAGCATCTAGAGAGGTGAG 591
QY 181 TCAGCTCTCTTTTCAACAGAAATCTAAATTGAAAGTCAGCTTTGATGAGAAAG 240
DB 592 ATGCTTTACTTAAATGATAGTAACTTTCCAAACAGAAAGCAATTCGAGTCC 651
QY 241 ATATTATTTTAACTAAAGACTACAGCAAAAGGTTACTGAAAGTTTAAAGACCCGG 300
DB 652 ATCTTTTCTGTACAAAGCTAGAAAGAAAGGTCAACAGAGTTGTCAAGTGAATCTAA 711
QY 301 AATGAGAGAGATGTAGAGTAAACGATCACTTTAAACAATGAACTTCCACCTACACCA 360
DB 712 AAGAGTGAACATAAAGATGACTATACCTGTAAGAGGAGAGCTGCATCAAGTTCTCC 771
QY 361 ACTTGTTCAGTTCATATATATTTTTCAGAGGCTTTGAAAATCATGAAATTTGCA 420
DB 772 GTGTGATCAGAGTCTTCAATATCATCTTCAAGAAAGATCTCAAAAAGTGTCAATGAC 831
QY 421 CAATTTGAGCAAAATTTATTAACCAATGACCCCAATGATATTTCCAAAGTACAGGTG 480
DB 832 CAATTTGAGCAAACTTCTATATCTTCAAGAGCCATGAAATTTCCCAAGCAAAATTA 891
QY 481 GTGATTTGGCTGGCTTCACTACTTCCATCTTCAGTATGAAAGAACGATCATGCTGTC 540
DB 892 TCCCTTTGGCTGGCTTCCATCTTCTGTGATATTTTGAAGAGAGTCTGTTTGT 951
QY 541 ACTGAGCTTACCCATTAAGTCTTGAAGTGAAGCTTTTGGATTTTCAATGTTCACTTT 600
DB 952 GGTGATGTGATTAAGAGTCTTCCGAAATGAGAGCGTTCTGGAATTCATGACTGCTTC 1011
QY 601 TATCATCAGACAGAAAGCAATAATTTCAAGAACTTTCCAAAGAACTAAATAGATTTA 660
DB 1012 TGTCAAGAACTGTGCTTCTGTTTCCACCAAGCTGTGAGAGAGAGTAAATAGGCTTC 1071

```

```

QY 661 GTTGTCTTACCAAGATTAACAATAAGACATACAGAGTGAATGATTTGACTGGACCG 720
DB 1072 ATTTGCTTTACAGATACATTAACGAACCTTCTCATTTGATGACATTTGACTGGACG 1131
QY 721 AATCCCAAGAGCACTTTTAAAGAAACCGAGCTCTGAAATGACCTTTCTTGAATATCTG 780
DB 1132 AAGCCACACACACTTTTCAAGAGCGGATGACACCGGATCACTATGTGATTAATCAC 1191
QY 781 AGAAGCAATTAACAACAAAGATCAACCACTTGAAGAGCGCTGTGGTGCACGACGCC 840
DB 1192 AAGCAGCAATTAAGATTAATCTGATGAGACTGAATCACCACATGCTTTGATGCTGTTA 1251
QY 841 AAGAGAGCGCGGCGCTGAGGAGGACATCCAGAGGCGCTGCATGCTCATCTTGAAGCTG 900
DB 1252 AAG--AAGAGAGAAATGACAAACAGTGAAGCTGACGCTGCCACCTGATACCTGAGCTC 1308
QY 901 TGTATCTTACAGTCTTACTGATTAATACCGTAATGATTTTAAAGGTGAAAGACTTA 960
DB 1309 TGTCTTTTAAACAGGCTGATCTGACAGGCAATCTGATTTTCAAGTGAAGGCTGTG 1368
QY 961 GCCGTTATACAAAGACTTAACTCCAGAGCAAGCGAGGAGTGAAGTGGACGACTGAT 1020
DB 1369 GCTGAAAGACACGCTGATGCTCCTTCAAGCGCGGACAGACGCTGGCCAGGCTTGAGAC 1428
QY 1021 TACATTCATTAATAACGATATATGTTCAAAGGAGCTTCAGACCTGGGCTTGAAGCTTGA 1080
DB 1429 AACATCCAGAGAAATACCAATGCTGCTTTGAAGTGAAGACCTGGGAGCTGCAATTTGGA 1488
QY 1081 TCCACTTACTGCTCTTCAAGAGAAATTTTGAACAGAAAGATTCACCAAGGTGGA 1140
DB 1489 AGCCAGATA---TCTGTGACTGGCGGATTTGGCTTGAAGAAATTAATTAATGCAAGAC 1545
QY 1141 AAAACATTTGATTAACAATCAATCAATTTGCAAGTGTCTCCAAAGAAACAAGAGTGA 1200
DB 1546 CACATAT---GTCAACCTGTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1602
QY 1201 TTAATTAAGTTTAAAGCAGATTAATCTGAGCTGTTGATCTATACCGGAAGAAATTAAGA 1260
DB 1603 ATTTTAAAGCAGAGCTTTGAATACCTGCTGATTTTATGATGAGCAGAGACTGAATAT 1662
QY 1261 GCAGCAATTCATTTGATTAACAAATCTATTTAAAGTTAAACACAGCCATGGGATGCA 1320
DB 1663 GTTCCGAGAGCTTTCTGAATCTGCTGAGAGAGGTTGAGAGTTCATGGATTTAATGTG 1722
QY 1321 AGAAAGCAATTAAGTGAAGTGAAGTGAAGAGACTGAAGCTTAAAGAGTCTTAAG 1380
DB 1723 GACTACCCCAAAATCTAAAGATTAAGAGAAATCCAGCTGCAATTTGTAAGGATATACG 1782
QY 1381 CAAAAGTCAACAGACACACCAAGATGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1440
DB 1783 CAATATGATCTGATATGATGTTGCTGATGATGATGATGATGATGATGATGATGATGATG 1842
QY 1441 AAATTCATCTTAAATTAATACTGTGTGAAGATTTGACCTTCAAGCTCAAGTGTGTG 1500
DB 1843 TATATGATCTCAATTAATAATTTGAGCTGCACTGCACTGCACTGCACTGCACTGCACTG 1902
QY 1501 GTGCGCCGAACCTTGAAGCAACAGCAACCTGTATGCGCATGCTTCTCAAAAATTTGCCCT 1560
DB 1903 CTGCTGCGACCTTGAATTAACAGGCGATGATGATGATGATGATGATGATGATGATGATG 1962
QY 1561 CAGATGAATGCAAGATGAGAGAGAGCTTCTGAGAGGTGAGCAATCCCGCGAAGCTGTG 1620
DB 1963 CAGATGAATGCAAGCTGAGAGAGAGCTGAGAGGTGAGCAATCCCGCGAAGCTGTG 2022
QY 1621 ATGATCGTGGCATTCATTTGTTTACATGACATGACAGCTGGGCGAGGTCAATGCGAGGA 1680
DB 2023 ATGATCGTGGCATTCATTTGTTTACATGACATGACAGCTGGGCGAGGTGATGATGATG 2082
QY 1681 TTTGTTGCGCAGATCAATGAAGAGATGACCCGCTGTGTTCAAGCTGATATTTTCAGAT 1740
DB 2083 TGCCTGCGCAGGTGTTAAACCCAGATCAACAGGAGGTTTTCGCGCTGATCTTCAAGAGA 2142
QY 1741 AAGAGCAGAGAGCTGTGATGAGTGGCTCAAGATCTGCTGCAAGCGGCTTGAAGGAGCTTGG 1800

```

Db	2143	ACAAATGACTGATGTTTGGACAGATTGCTGAAAGTTTTCATGATGTGAGACACTCAACAAATGG	2202
Qy	1801	AATAGCTGCATATGATGATCATATGCCAGCCGATCATCGTGTACCGGATAGGGGTAGAGAC	1866
Db	2203	TACAAAGTACATCATGATTTGGCCAGCACGGAATATGTGATCCGCTGTGTAGGGAT	2262
Qy	1861	GGCCAGCTGAAAAACACCTGGTGAACCTACAGATGCGACAGTTTGGATTGCTTAAATCC	1920
Db	2263	GGTACGCTGAAAAACCTATTATGATATTAATATCCACAGCTGTGAGACAGTGTGGACAA	2322
Qy	1921	ATTGTGATAGGTTTACAACTTACATGACTTAAACGGTATTTGTGTGAAAAAAAGAGTACACC	1988
Db	2323	TCCAGCTCAAAATACAGCTCAAGACTGTGCGTGTATGTGGTACAGAAAGATGACATGCCA	2382
Qy	1981	AGATTTTGTGTCAGTGTGAGGAAGAATTGAGAAATCCACTTCCGTGAACATTATGAT	2040
Db	2383	CGATTCTTTACCGAATATACACCGCACTGTACGAACCCCCCACTTGGACATGTTGTGAT	2442
Qy	2041	GTAGAGGTTTACAGACCAAGAAATGGTATGACTTTTATCTGAGCCAGGCTGTGAAGT	2100
Db	2443	TCAGAGCAACACGTAAAGAAATGGTATGACTTTTATCTGATCAGCCAGAGTGGCTGGCCG	2502
Qy	2101	GGTATGTTTCTCCCAACATTTACATATCTATCTATGACAAACAGGGGCTGAAGCAGAC	2160
Db	2503	GGAACTGTATGTCTTACCTACTATATATCTATATCTATATGATGACAAAGGCTTGAAGCCGAC	2562
Qy	2161	CACATACAGGCTTGAACCTTACAGGCTGTGCCACATCTATTTTCAACTGGCCAGTGTCAAT	2220
Db	2563	CATATGACGAACTTACATGATCAAAATGTGTGCCACCTGTACTAACTGGCCGGGCAATATGC	2622
Qy	2221	CGTGTCTCTGCTCCTTGGCAGTACGCCCAACAGCTGCTTTTCTGTGTGGCCAGTAT	2280
Db	2623	AGTGTCCACAGACCAATGTACAGATGTCTCAACAGCTGACCTTTCTGTGTGCCAAGAACAT	2682
Qy	2281	CACAGAGGCCAAATCTGTCACTGTCAAAACGCTTACTTACACTCTAA	2338
Db	2683	CATTAAGAACCCAGCTGTGAAATTTAGCCAAACCATCTCTTCTTACCTGTGA	2730
RESULT 15			
AX834401		2456 bp	DNA
LOCUS	AX834401		linear
DEFINITION	Sequence 1525 from Patent EP1347046.		PAT 15-DEC-2003
ACCESSION	AX834401		
VERSION	AX834401.1		GI:39920536
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 Isoga,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,U.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuko,Y.		
JOURNAL	Full-length cDNA sequences		
FEATURES	Patent: EP 1347046-A 1525-24-SRP-2003; Research Association for Biotechnology (JP)		
source	Location/Qualifiers		
	1..2456		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	26.5%;	Score 616.8;	DB 6;
Best Local Similarity	57.5%;	Pred. No. 5,2e-155;	Length 2456;
Matches 1170;	Conservative	0; Mismatches 857;	Indels 9; Gaps 3
Qy	293	AATCCCGAATGAGAGATGAGATTAACATGATCACTTTAAACAATGAATTCACACTTA	352
Db	2	AATCTAAAGGTGAGACTATTAAGATGATCTTACCCCTGAAGAGAGGCTCATCA	61

QY	355	CATCACCAACTGTTGTTGGAGTCTCATATAATTTATTTTCAGGAGGCTTTGAAATCATGA	412
Db	62	GTCTCCCGGTGATCCAGGCTTTCAATATCATCTTCGAAAAGATCTCCAAAAGTTGG	121
QY	413	ATTTCGACCAATTGGACGAAATTATTAATTAACCCAAATGACCCAAATGATATTCGAAATC	472
Db	122	CCATGTGACCAAAATTTGGACGGAACTTCTATATATCTTCAAGGCCAAATGGAAATTTCCGACG	181
QY	473	ACAGGTGATGATTTGGGCTGGGCTCAACATCTTCCATCTCTCAGTATGTAAGAACAGATCA	532
Db	182	ACAAATTAATCCCTTTGGCTCGGGTTGGCAATTTCTGTGATATTTTGAAGAGAGCTCC	241
QY	533	TGCTGTGACTGACGTTAGGCCATPAAGTCTTTGAAATGAGACTGTTTGGATTTCAATG	592
Db	242	TGTTTATGTGTGATGTGATTAACAAAGTCTCCGGAATGAGAGGTTCTGGAAATTCATGA	301
QY	593	TCAACTTTATTCATCAGACAGAGAAACATAAATTTCAAGAACAGTTTCCAAAACATA	652
Db	302	CTGCTCTCTGTCAAAAGACTGGCTGTGCTGTTCACCCAGAGGTGTGAGAACAGATTA	361
QY	653	TAGGTTATGTTGTTCTTACAAAGTATPAACAATAAGACATCAGAGTGTGATATTAAGCT	712
Db	362	TAGGGCTCATGTGCTCTTACAAATACATPAACGAACCTATCCATTTGAAGAACATTA	421
QY	713	GGAGACGAATATCCCAAGACACTTTAAAGAACCGACGAGCTCTGAAGTCAAGCTCTTA	772
Db	422	GGTACGTGAAGCCCAACACACCTTTTAAAGGGAGATGGCAACGAGATCACTATGTGG	481
QY	773	AATATCTACAGAGCAATPAACAACCAAGATCAACGACCTTGAAGCAGCTGTGTGTCA	832
Db	482	ATTATCTACAGACGACGATATGATATTACTGTATCGACCTGAATCAACCCATGCTTTA	541
QY	833	GCCAGCCCAAGAGAGCGGGGCTCTGGGGGACACTGCGAGGGCTGTGCATGCTCATC	892
Db	542	GTCGTTTAAAG--AAGAAAGAAATGACAAAGTGAAGCTCAGCTGAGCCACCTGAATAC	598
QY	893	CTGAGCTCTGTCTATCTTACAGGTCTAACTGTATTAATATGCTAATGATTTTAACGTATGA	952
Db	599	CTGAGCTCTGTCTTCTTAAACGGGCTGTGACTGACGGAACATCTGATTTTCCAGCTGAATGA	658
QY	953	AAGACTTAGCCGCTCATPAACAAGCTAACTACGACGACAAAGCAGCGTGAATGGAGACAC	1012
Db	659	AGGCTGTGCTGTAAGAAAGACAGTCTCAGTCTTGAAGCCCGGACAGCGCTGGCCAGGC	718
QY	1013	TCATTTGATTCATTCATPAAAAACATATATGTTCAAAGGAGCTTCGAGACTGGGGGTTGA	1072
Db	719	TTTGAGACAAATCCAGAGGAATACCAATGCTGCTTTGAATCTAGAGACCTGGGGACCTGC	778
QY	1073	GCTTTGATTCCAACTACTGCTCTTCAGAGAAATTTTGGCAACGAAAAGATTCACC	1132
Db	779	ATTTTGGAGGACATATA--TCTGTGACTGGCCGAAATGTGCTCTTCGAAAATTAATPAA	835
QY	1133	AAGGTGAAAAACATTTGATTTACATCAACAATTTGACATTTGGTCCAAAGAAACAAG	1192
Db	836	TGCAAGACCAATAT--GTCAACCTGTGTGCTGTGACTGTGCTCCAGGATTTTGCAA	892
QY	1193	GTGACCACTPAATTAAGTGTAAAGCCACTAGATACTGGCTGTGATCTATAACGGAGAA	1252
Db	893	CTTGCAAAATTTTAATGACACAGCTTTGAATACCTGGTTGATTTTAATGTGACACAGAA	952
QY	1253	ATTATGAAGCAGCCCAATTCATTTGATTAACAAATCTATTTAAAGTATCACACGACATGGCA	1312
Db	953	CTGATATATGTTGCCAGAGCTTCTGAAGCTGCTTGAAGAGGTTACAGCTTCCATGGAT	1012
QY	1313	TGCAATATAGAAAAGCAATATATGATTTGAAGTGTATGACAGAACTGAAGCTTAATTAAG	1372
Db	1013	TTTAATGTGACTACCCCAAAATCATATAAAGTACAGAAAATCCAGCTCATTTGTTAAG	1072
QY	1373	TCTTAACGCAAAAGGTCAACAGACACCCCAATATGTTGTCTGTCTGTGCAAGTATAC	1433
Db	1073	CTATACACCAATATGTGATCTGTATGTTCAAGTGTGATATGTGACTTCTGCTTTATATC	1133

Wed May 12 09:50:20 2004

us-10-043-774b-1.rge

Page 24

QY	1433	GGAAAGCAAAATACAGATGCTATTAAATAATCCGTGTGAACAAGATTGGCCTAACCCCAAGTC	1492
Db	1133	AGAAAGCCCTATTATGATTTCAATTAATAATATTAGACTCAAGTCCCAAGTCCCAAGCC	1192
QY	1493	AGTGTGTGTGTGGCCCGCAACCTTTAGCGAAACAGCAAACTGTCAATGGCCATTGCTACAAAGA	1552
Db	1193	AATGTGTGCTTGTCTGGAGCCTTGATTAATAACAGGGCAGATGATGATGATGCGCCACAAAGA	1252
QY	1553	TTGGCCCTACAGATGAATCTGCAAGATNGGAGAGAGCTCTGGAGGGTGGAGACATCCCCCTGA	1612
Db	1253	TCCGTATGTCAATGACTTGGCAAGCTCGAGGGCCGAGCTGTGGCTGTGGAAATACCTTTAA	1312
QY	1613	AGCTCGTATATGCTGTGGCATTCGATTGTATTACATGACATGACAGCTGGGCGAGGTCAA	1672
Db	1313	AGTCCCTGATGTGTGTGTCTGGATTTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGG	1372
QY	1673	TCGCAGATTTGTGTGCCAGCATCAATGAAGGATGACCCCGTGGTCTTCAGAGCTGCATAT	1732
Db	1373	TTGTGTGATGTGGTGGCCAGTGTAAACCCCAAAATCACAGGTGGTTCCTCCGCTGTATCC	1432
QY	1733	TTTCAGATAGAGGACAGGAGGCTGGTGAATAGGAGCTCAAGTCTGCCTCGAAACGGCTCTGA	1792
Db	1433	TTTCAGAGAACATGACTGATGTGTGAGATTTGCTTGAAGTTTTCATATGACTGGACATCA	1492
QY	1793	GGGCTTGGAATAGCTGCATATGATACATGCCCAAGCCGGATCATGCTGTACCGCATGGCG	1852
Db	1493	ACAATGTGTAACAATGACATCATGATTTTGGCAGCAAGGATTAATGTGTACCGTCTGTGTG	1552
QY	1853	TAGAGAGAGGCGCAGCTGAAAACACATCGGTGAACCTACGAAGTGCACAGTTTGTGATTTTC	1912
Db	1553	TAGGGGATGTGTAGCTGTGAATACTTATTGATATGAAGTCCCAAGCTGTGTAGAGACTG	1612
QY	1913	TAAATTCATTGGTGTAGAGTTACAAACCTTAGCTAACGGTATATGTGTGTGAAGAAAGAG	1972
Db	1613	TGGCAGAAATCCAGCTCAAAATACCACTCAAGCTCTCGGTGATGTGTGTCAAGAAAGAT	1672
QY	1973	TGAACACACAGATTTTGTGTGCTGCTGAGAGAAAGCTTCAGAAATCACTTCCTGTGGAACG	2032
Db	1673	GCATGCCAGATTTCTTACCGAAATGAACCGCAGTGTGAAGAACCCCCCACTTGGCATGTG	1732
QY	2033	TTATTGATGTAGAGTTACCAAGCCAGATGGTATGACTTTTATTCGTGAGCCAGGCTG	2092
Db	1733	TTGTGTGATTCAGAAAGAACAGTAAAGATGTATGACTTTTATCTGATCAGCCAGGTGG	1792
QY	2093	TGAAGATGTGTGTGTCTTCTCCACACATTAAGATCATCTATATGAACAAGGGCCCTGA	2152
Db	1793	CTGTCCGGGGAACGTGTACTTACTTACTATTAATGTACTATATATGACAAGGCTTGA	1852
QY	2153	AGCGAGACACATACAGCGCTTGACCTTCAAGCTGTGCCACATCTATTTACAAGTGGCAG	2212
Db	1853	AGCCCGAACATATGAGAGACTTACATTAATATGTGGCACTGTATCTRAACAATGCCGG	1912
QY	2213	GTGTCAATTCGTGTCTCTCTCTTGGCAAGTACGCCCAAGCTGGCTTTTCTGTGTGGCC	2272
Db	1913	GGATATGTAGTGTCTCCAGACACATGTCAATATATGTCTACAAGGTGAGCTTGTGTGGCAC	1972
QY	2273	AGAGATTACAGAGAGCAAAATCTGTCACTGTCAAAACCGCTTACTACTCTTAA	2332
Db	1973	AAAGCACTTAAATGAAGAACCAAGTCTGGAAATTAGCCACCAATCTCTTCTTACCTGTGA	2032

Search completed: May 8, 2004, 14:09:03
Job time : 9063 secs

Wed May 12 09:50:21 2004

us-10-043-774b-1.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 11:04:37 ; Search time 5969 Seconds
(without alignments)
11646.701 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328
Sequence: 1 atgaccttgcgtgcaacac.....acgccttactactctaa 2328

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthm:*
3: em_estmu:*
4: em_estov:*
5: em_estpl:*
6: em_estro:*
7: em_hic:*
8: gb_est1:*
9: gb_est2:*
10: gb_est3:*
11: gb_est4:*
12: gb_est5:*
13: em_estfun:*
14: em_estom:*
15: em_ges_hum:*
16: em_ges_inv:*
17: em_ges_pin:*
18: em_ges_vrt:*
19: em_ges_fun:*
20: em_ges_mam:*
21: em_ges_mus:*
22: em_ges_rnd:*
23: em_ges_vrt:*
24: em_ges_vrt:*
25: em_ges_vrt:*
26: em_ges_vrt:*
27: em_ges_vrt:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2316.2	99.5	2586	29	AY410583 Homo sapi
2	1967.2	84.5	2586	29	AY410584 Pan trogl
3	1684.6	72.4	2589	29	AY410585 Mus muscu
4	575.6	24.7	874	10	BF666315

5	515.2	22.1	577	10	BF080918
6	513.4	22.1	527	12	BM751055
7	502.4	21.6	1495	10	BF646941
8	492	21.1	705	10	BF246708
9	471.2	20.2	533	10	BF080923
10	459	19.7	3559	11	AK030116
11	441.4	19.0	462	9	AA969938
12	439.8	18.8	818	13	BF773137
13	433.2	18.6	642	12	BU854142
14	415.2	17.8	565	12	BG101107
15	413	17.7	413	9	AL041473
16	405.6	17.4	779	12	BI560710
17	400.4	17.2	664	13	BU234973
18	390.6	16.8	1092	12	BI463482
19	388.8	16.7	420	13	BO374910
20	387	16.6	587	9	AA430311
21	373	16.0	935	10	BF247005
22	370.8	15.9	608	13	BU230168
23	365.4	15.7	767	13	BU232120
24	353.4	15.2	523	14	CK137141
25	350.6	15.1	875	14	CA471292
26	347	14.9	882	12	BI459464
27	324.6	13.9	715	12	BF721550
28	307.8	13.2	666	12	BI463897
29	305.8	13.1	622	9	AL705072
30	301	12.9	731	12	BI458664
31	296.8	12.7	640	9	AL705301
32	296	12.7	833	13	BU275338
33	285.6	12.3	767	14	CB592670
34	285.2	12.3	609	13	BO481304
35	275.8	11.8	559	12	BM728018
36	272	11.7	807	13	BU160046
37	265	11.4	617	12	BO400023
38	264.8	11.4	512	12	BM680592
39	258.4	11.1	477	12	BM539361
40	257.6	11.1	2619	29	AY407009
41	249	10.7	718	13	BO435882
42	245.2	10.5	568	13	BO074332
43	243.4	10.5	785	12	BG221179
44	242.2	10.4	378	10	BF764928
45	241.2	10.4	534	12	BI672860

ALIGNMENTS

RESULT 1
LOCUS AY410583
DEFINITION Homo sapiens HCM3939 gene, VIRUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY410583
VERSION AY410583.1 GI:39766551
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2586)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtinsky,J.J.,
Adams,M.D. and Carcilli,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2586)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtinsky,J.J.,
Adams,M.D. and Carcilli,M.
TITLE Direct Submission

Wed May 12 09:50:21 2004

us-10-043-774b-1.rst

Page 2

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES Location/Qualifiers

```
source      1. .2586
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
gene        <1. .2586
            /locus_tag="HCM339"
```

ORIGIN

Query Match	99.5%	Score 2316.2	DB 29	Length 2586
Best Local Similarity	99.9%	Pred. No. 0		
Matches 2318; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY	8	TTGGTGTGAACAACAAGCGAAGCTTAGACCATGTTAAAGATCAAAAAACAGTTCCTTCAG	67
Db	266	TTGGTGTGAATAACAAGCGAAGCACTAGACCATGTTAAAGATCAAAAAACAGTTCCTTCAG	325
QY	68	GCATTATATGTAAGGTTAAGCACTAACCATTTCTCCGGCTGCACATCCCGTCCCAAGTGGGACT	127
Db	326	GCATTATATGTAAGGTTAAGCACTAACCATTTCTCCGGCTGCACATCCCGTCCCAAGTGGGACT	385
QY	128	TATATCACTATCACATTGACTATTAACCCACTGATGGAAGCCAGAAGATCTCCGTTCACTC	187
Db	386	TATATCACTATCACATTGACTATTAACCCACTGATGGAAGCCAGAAGATCTCCGTTCACTC	445
QY	188	TTCTTTTTCACAACGAAGATCTAATTTGGAAGTGTCACTGCTTTTGTGATGGAACGATATAT	247
Db	446	TTCTTTTTCACAACGAAGATCTAATTTGGAAGTGTCACTGCTTTTGTGATGGAACGATATAT	505
QY	248	TTTATACCTAATAAGACTACAGCAAAAAGTTACTGAAGTTTATGTAAGACCCGAGATGAG	307
Db	506	TTTATACCTAATAAGACTACAGCAAAAAGTTACTGAAGTTTATGTAAGACCCGAGATGAG	565
QY	308	AGGATGTGAGATTAACGATCACTTTAACAAATGAACTTCCACTACATACCAACCTGTT	367
Db	566	AGGATGTGAGATTAACGATCACTTTAACAAATGAACTTCCACTACATACCAACCTGTT	625
QY	368	TGCATGTTCTATTAATTTATTTTTCAGGAGGCTTTGAAATATCATGATTTTGCAACAAATTG	427
Db	626	TGCATGTTCTATTAATTTATTTTTCAGGAGGCTTTGAAATATCATGATTTTGCAACAAATTG	685
QY	428	GACGAATTTATTAATACCCAATGACCCCAATTTGATTTCCAAGTCAACAGTTGTGATTT	487
Db	686	GACGAATTTATTAATACCCAATGACCCCAATTTGATTTCCAAGTCAACAGTTGTGATTT	745
QY	488	GGCCTGGCTTCACTACTTCCTCCATCGATATGAAAAACAGATCATGCTCTGCACGTACG	547
Db	746	GGCCTGGCTTCACTACTTCCTCCATCGATATGAAAAACAGATCATGCTCTGCACGTACG	805
QY	548	TTAGCGATTAAGTCCCTGGAAGTGAACGTGTTTGGATTTCAATGTTCAACTTTATCATC	607
Db	806	TTAGCGATTAAGTCCCTGGAAGTGAACGTGTTTGGATTTCAATGTTCAACTTTATCATC	865
QY	608	AGACGAAGAACATTAATTTTCAAGAACAGTTTCCAAAGAACTAATAGGTTTAGTTGTTTC	667
Db	866	AGACGAAGAACATTAATTTTCAAGAACAGTTTCCAAAGAACTAATAGGTTTAGTTGTTTC	925
QY	668	TTACCAAGATTAACAATTAAGACATCAAGAGTGGATGATTAATGACCTGGGACCAAGATCCCA	727
Db	926	TTACCAAGATTAACAATTAAGACATCAAGAGTGGATGATTAATGACCTGGGACCAAGATCCCA	985
QY	728	AGAGCACTTTTAAAGAAAGCCGACGGCTCTGAAGTCAGCTTTTGAATATCTACAGGAAGC	787
Db	986	AGAGCACTTTTAAAGAAAGCCGACGGCTCTGAAGTCAGCTTTTGAATATCTACAGGAAGC	1045
QY	788	AATACCAACCAAGATCAACCGACTTGAAAGACCGCTCTGGTTCAGCCAGGCCAAGAAGAA	847
Db	1046	AATACCAACCAAGATCAACCGACTTGAAAGACCGCTCTGGTTCAGCCAGGCCAAGAAGAA	1105

QY	848	GCGCGGCGCCTCGGGGGGACCTGCGCAGGCGCTGCGATGCTCATTTCCGAGACCTCGCATC	907
Db	1106	GCGCGGCGCCTCGGGGGGACCTGCGCAGGCGCTGCGATGCTCATTTCCGAGACCTCGCATC	1165
QY	908	TTACAGGCTCTACGTATTAATGCGTATATTTTAACTGATGAAAGACTTAGCCGCTTC	967
Db	1166	TTACAGGCTCTACGTATTAATGCGTATATTTTAACTGATGAAAGACTTAGCCGCTTC	1225
QY	968	ATAACAGACTPACTCCAGACAAAGCGCTGTAAGTGGGACGACTCATTTGATTACATTC	1027
Db	1226	ATAACAGACTPACTCCAGACAAAGCGCTGTAAGTGGGACGACTCATTTGATTACATTC	1285
QY	1028	ATATAAACGATANTGTTCAAAAGGAGCTTGAGCTGGGGTTTGAGGCTTGATTCCAACT	1087
Db	1286	ATATAAACGATANTGTTCAAAAGGAGCTTGAGACTGGGGTTTGAGGCTTGATTCCAACT	1345
QY	1088	TACTGTCCCTCTCAGAGAGATTTTTCGAAACAGAAAAGATTCAACAAAGTGGAAAAACAT	1147
Db	1346	TACTGTCCCTCTCAGAGAGATTTTTCGAAACAGAAAAGATTCAACAAAGTGGAAAAACAT	1405
QY	1148	TTGATTTCATCCCAATTTTGCAGATTGTCGCAAGAAACAAGAGTGCACATTAATTA	1207
Db	1406	TTGATTTCATCCCAATTTTGCAGATTGTCGCAAGAAACAAGAGTGCACATTAATTA	1465
QY	1208	GTTGTTAAGCACTAGATPACTGCTGTGTGATTCTTAACGCAAGAAATTAATGAACACCA	1267
Db	1466	GTTGTTAAGCACTAGATPACTGCTGTGTGATTCTTAACGCAAGAAATTAATGAACACCA	1525
QY	1268	ATTGATTGATPACAAATCTTATTTTAAAGTTACACGACCATGGGCAATCCAAATGGAAG	1327
Db	1526	ATTGATTGATPACAAATCTTATTTTAAAGTTACACGACCATGGGCAATCCAAATGGAAG	1585
QY	1328	CAATPATGATTGAAGTGGATGACAGAACTGAAAGCTTCTTAAGAGTCTTACGCAAAAG	1387
Db	1586	CAATPATGATTGAAGTGGATGACAGAACTGAAAGCTTCTTAAGAGTCTTACGCAAAAG	1645
QY	1388	TCACAGACGACCCGAGATAGTTGTCGTCTGTTGTCAAGTAAATCGGAAGACAAATACG	1447
Db	1646	TCACAGACGACCCGAGATAGTTGTCGTCTGTTGTCAAGTAAATCGGAAGACAAATACG	1705
QY	1448	ATGCTATTTAAAAATTAACCTGTTGACAAATTGCCCTACCCCAAGTCAGTGTGGTGGCCC	1507
Db	1706	ATGCTATTTAAAAATTAACCTGTTGACAAATTGCCCTACCCCAAGTCAGTGTGGTGGCCC	1765
QY	1508	GAACTTATGGCAAAACAGCAAACTGTGTCATGGCCATTGCTACAAAGATTGCCCTACATGA	1567
Db	1766	GAACTTATGGCAAAACAGCAAACTGTGTCATGGCCATTGCTACAAAGATTGCCCTACATGA	1825
QY	1568	ACTGCAAGATGGGAGAGAGCTCTGGAGGGTGAACATCCCTGTAAGCTGTATGTATCG	1627
Db	1826	ACTGCAAGATGGGAGAGAGCTCTGGAGGGTGAACATCCCTGTAAGCTGTATGTATCG	1885
QY	1628	TTGCAATCGATTTGTACCATGACATGACAGCTGGCGGAGGTCAATCGACGATTTGTTG	1687
Db	1886	TTGCAATCGATTTGTACCATGACATGACAGCTGGCGGAGGTCAATCGACGATTTGTTG	1945
QY	1688	CCAGCATCAATGAAGGAGTGAACCCGCTGTTCTACGCTGCAATTTTCAGATGAGAGAC	1747
Db	1946	CCAGCATCAATGAAGGAGTGAACCCGCTGTTCTACGCTGCAATTTTCAGATGAGAGAC	2005
QY	1748	AGAGAGCTGTGATGAGGCTCAAACTCTGCTGCACAGGCGCTCTGAGGCTTTGGAATAGCT	1807
Db	2006	AGAGAGCTGTGATGAGGCTCAAACTCTGCTGCACAGGCGCTCTGAGGCTTTGGAATAGCT	2065
QY	1808	GCAATGACTACATGCCCGGATCATCGTGTACCGTGAATGGGTTAGGAACGGCCAGC	1867
Db	2066	GCAATGACTACATGCCCGGATCATCGTGTACCGGGAATGGGTTAGGAACGGCCAGC	2125
QY	1868	TGAATAACCTGATGAACACAGATGCGACAGTTTGTGATTTGATTTCTAAATTCATTTGTA	1927
Db	2126	TGAATAACCTGATGAACACAGATGCGACAGTTTGTGATTTGATTTCTAAATTCATTTGTA	2185
QY	1928	GAGGTACACCTTAGACTTAAACGATATTTGTGTGAAGAAAAAGTGAACACAGATTTT	1987

Db 2186 GAGGTACCAACCTTACGATTAACGATTAATTTGATGAGAAAGAAAGTGAACACCAATTTT 2245

Qy 1988 TTGCTCAGCTTGAGAGAAAGCTTCAAGAAATCCATTCCTGGAACAGTTATATGATGAGG 2047

Db 2246 TTGCTCAGCTTGAGAGAAAGCTTCAAGAAATCCATTCCTGGAACAGTTATATGATGAGG 2305

Qy 2048 TTACCAAGACCAAAATGATGATGATTTTATTCGTAAGCAGGCTGTGAGAGTGTAGTG 2107

Db 2306 TTACCAAGACCAAAATGATGATGATTTTATTCGTAAGCAGGCTGTGAGAGTGTAGTG 2365

Qy 2108 TTCTCCACACATTAACATGTCATCTATGACAAAGGCGCTGAGGCAAGCAATAC 2167

Db 2366 TTCTCCACACATTAACATGTCATCTATGACAAAGGCGCTGAGGCAAGCAATAC 2425

Qy 2168 AGCGCTTGACCTTAACAGCTGTGCAATCTATTAACATGCGCAGAGTGTATGCTGTC 2227

Db 2426 AGCGCTTGACCTTAACAGCTGTGCAATCTATTAACATGCGCAGAGTGTATGCTGTC 2485

Qy 2228 CTGCTCCTTGACAGTACGCGCAAGAGCTGTTTCTTGTGTCAGAGATTCACAGAG 2287

Db 2486 CTGCTCCTTGACAGTACGCGCAAGAGCTGTTTCTTGTGTCAGAGATTCACAGAG 2545

Qy 2288 AGCAAAATGTGCTACTGTCAAAACCGCTTTACTACTCTAA 2328

Db 2546 AGCAAAATGTGCTACTGTCAAAACCGCTTTACTACTCTAA 2586

RESULT 2
AY410584 2586 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HCM9399 gene, VIRUAL TRANSCRIPT, partial sequence,
AY410584
ACCESSION AY410584
VERSION AY410584.1 GI:39766552
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
AUTHORS 1 (bases 1 to 2586)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civeallo, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2586)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civeallo, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Substitution
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. 2586
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..2586
/locus_tag="HCM9399"

ORIGIN
gene

Query Match 84.5%; Score 1967.2; DB 29; Length 2586;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1972; Conservative 0; Mismatches 349; Indels 0; Gaps 0;

Qy 8 TTGGTGTGAACACAGGCAAGCACTTACCACTGTATTAAGATCAAAAACAGGTTCTTCAG 67

Db 266 TTGGTGTGAATACAGGCAAGCACTTACCACTGTATTAAGATCAAAAACAGGTTCTTCAG 325

Qy 68 GCATTATAGTAAAGTTAGAGCTTAACCACTTTCCGGCTGACATCCGCTCCCAATGGGCTT 127

Db 326 NNN 385

Qy 128 TATATCAGTATCACTTATGATTAACCACTGATGAGAGCAGAGAATCTCCGTTGAGCTC 187

Db 386 TATATCAGTATCACTTATGATTAACCACTGATGAGAGCAGAGAATCTCCGTTGAGCTC 445

Qy 188 TTCTTTTCAACGAAAGTCTAATTGAAAAGTGTATGCTTTGATGGAACGATATAT 247

Db 446 TTCTTTTCAACGAAAGTCTAATTGAAAAGTGTATGCTTTGATGGAACGATATAT 505

Qy 248 TTTTACTAATAAGACTAAGCAAAAGTTACTGAAGTTTATAGAACCCGAAATGAG 307

Db 506 NN 565

Qy 308 AGATGTAGAGATTAACGATCACTTAAACAATGAATCTCACTATCACTCACTTGT 367

Db 566 AGATGTAGAGATTAACGATCACTTAAACAATGAATCTCACTATCACTCACTTGT 625

Qy 368 TGCAATCTTAATATATATTTTCAAGAGGCTTTGAAAATCATGAATTTGCAAAATG 427

Db 626 TNNNNNTCTAATATATATTTTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 685

Qy 428 GACGAATTAATTAACCAAAATGACCAATGATATTCAGAGTCAAGGTTGTGATTT 487

Db 686 NN 745

Qy 488 GGCTGGCTTCACTACTTCACTCTTCAATGATGATGATGATGATGATGATGATGATGATG 547

Db 746 GGCTGGCTTCACTACTTCACTCTTCAATGATGATGATGATGATGATGATGATGATGATG 805

Qy 548 TTAGCCATTAAGTCTTCAAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTC 607

Db 806 TTAGCCATTAAGTCTTCAAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTC 865

Qy 608 AGACAGAAACATTAATTTCAAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTC 667

Db 866 AGACAGAAACATTAATTTCAAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTC 925

Qy 668 TTACCAAGTATTAACAATTAAGACATACAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTC 727

Db 926 TTACCAAGTATTAACAATTAAGACATACAGAGTGAAGTGTGATTTGATTTGATTTGATTTGATTC 985

Qy 728 AGACACCTTTAAGAAAGCGAGCGCTTGAAGTCACTTTGATTAATTAACAAGAAAGC 787

Db 986 AGACACCTTTAAGAAAGCGAGCGCTTGAAGTCACTTTGATTAATTAACAAGAAAGC 1045

Qy 788 AATTAACAAGAGATGACCGAGCTTGAAGAGCTGCTGTCAGCCAGCCAGAGAA 847

Db 1046 NN 1105

Qy 848 GCGCGGCGCTTGGGGGACATGCGCAGAGGCTGCGATGCTTCTGAGCTCTGCTATC 907

Db 1106 GCGCGGCGCTTGGGGGACATGCGCAGAGGCTGCGATGCTTCTGAGCTCTGCTATC 1165

Qy 908 TTACAGCTTCACTGATTAATTAAGTGAATTTTAAAGTGAAGAGCTTACCGCTTC 967

Db 1166 TTANNNGCTTAACGATTAATTAAGTGAATTTTAAAGTGAAGAGCTTACCGCTTC 1225

Qy 968 ATTAACAAGTAACTCAAGAGCAAGGAGCGTGAAGTGAAGCACTATGATTAATTC 1027

Db 1226 ATTAACAAGTAACTCAAGAGCAAGGAGCGTGAAGTGAAGCACTATGATTAATTC 1285

Qy 1028 ATTAACAAGTAACTCAAGAGCAAGGAGCGTGAAGTGAAGCACTATGATTAATTC 1087

Db 1286 ATTAACAAGTAACTCAAGAGCAAGGAGCGTGAAGTGAAGCACTATGATTAATTC 1345

Qy 1088 TACTGCTCTTCAAGAGATTTTGAAGCAAGAAAGATTCACCAAGTGAAGAAATCAT 1147

Db 1346 TACTGCTCTTCTCAGGAAAGATTTTGCAGAAAGAAAAGTTTCATCAAGTGTGAAAAACAN 1405
 Qy 1148 TTGATTAACAATCCCAATTTGACAGATTGTGTCAGAAAGAAACAAGAGTGCCATTAATTA 1207
 Db 1406 NTGATTAACAATCCCAATTTGACAGATTGTGTCAGAAAGAAACAAGAGTGCCATTAATTA 1465
 Qy 1208 GTGTTAAGCCACTGATTAAGTGTGTTGATCTTAAGAGAAAGAAATTAAGAGAGCC 1267
 Db 1466 GTGTTAAGCCACTGATTAAGTGTGTTGATCTTAAGAGAAAGAAATTAAGAGAGCC 1525
 Qy 1268 ATTCAATTGATAAATAATCTATTAAAGTTACACAGCCATGGGATGCAATGAGAAAG 1327
 Db 1526 ATTCAATTGATAAATAATCTATTAAAGTTACACAGCCATGGGATGCAATGAGAAAG 1585
 Qy 1328 CAATTAATGATTAAGTGTGATGACAGAACTGAAAGCTTAAGAGTCTTAAGAGAAAG 1387
 Db 1586 CAATTAATGATTAAGTGTGATGACAGAACTGAAAGCTTAAGAGTCTTAAGAGAAAG 1645
 Qy 1388 TCACAGCAGACACCCAGATAGTGTGTCTGTCTGTTGCAAGTAATCGGAAGACAAATACG 1447
 Db 1646 TCACAGCAGACACCCAGATAGTGTGTCTGTCTGTTGCAAGTAATCGGAAGACAAATACG 1705
 Qy 1448 ATGCTATTAAAAATACCTGTGTACAGATTGCCCTACCCAGTCAAGTGTGTGTGCCC 1507
 Db 1706 ATGCTATTAAAAATACCTGTGTACAGATTGCCCTACCCAGTCAAGTGTGTGTGCCC 1765
 Qy 1508 GAACCTTAGGCAACAGCAAACTGTCAATGCTGCTGCTCAATGCTCAAAAGATTGCCCTACAGATGA 1567
 Db 1766 GAACCTTAGGCAACAGCAAACTGTCAATGCTGCTGCTCAATGCTCAAAAGATTGCCCTACAGATGA 1825
 Qy 1568 ACTGCAAGATGAGAGAGAGAGCTCTTGAGAGGTGACATCCCTCTTAAGTCTGTATATCG 1627
 Db 1826 ACTGCAAGATGAGAGAGAGAGCTCTTGAGAGGTGACATCCCTCTTAAGTCTGTATATCG 1885
 Qy 1628 TTGGATGATTTGTTACCATGACATGACAGCTGGGCGGAGGTCAATGCAAGATTTGTG 1687
 Db 1886 TTGGATGATTTGTTACCATGACATGACAGCTGGGCGGAGGTCAATGCAAGATTTGTG 1945
 Qy 1688 CCAGCATCAATGAAAGAGATGACCCGCTGTTCTCAAGCTGCAATTTTCAGANTGAGAGAC 1747
 Db 1946 CCAGCATCAATGAAAGAGATGACCCGCTGTTCTCAAGCTGCAATTTTCAGANTGAGAGAC 2005
 Qy 1748 AGGAGCTGTGATGAGTGTCTCAAAAGTCTGCGAAGGCGCTCTGAGGCTTGAATAGCT 1807
 Db 2006 AGGAGCTGTGATGAGTGTCTCAAAAGTCTGCGAAGGCGCTCTGAGGCTTGAATAGCT 2065
 Qy 1808 GCAATGATGATCATGCGCCAGCCGATCATGCTGTATCCGCGATGGGTGAGAGACGCGCAC 1867
 Db 2066 GCAATGATGATCATGCGCCAGCCGATCATGCTGTATCCGCGATGGGTGAGAGACGCGCAC 2125
 Qy 1868 TGAAGAACTGTGTGATCAAGAGTCCACAGCTTTTGGATTTGTCTAAATTCATTTGTA 1927
 Db 2126 TGAAGAACTGTGTGATCAAGAGTCCACAGCTTTTGGATTTGTCTAAATTCATTTGTA 2185
 Qy 1928 GAGGTTACAACTCTGATCAATGATGATGTAAGTGTGTAAGAAAGATGAAACACAGATTT 1987
 Db 2186 GAGGTTACAACTCTGATCAATGATGATGTAAGTGTGTAAGAAAGATGAAACACAGATTT 2245
 Qy 1988 TTGCTCAGTCTGAGAGAGAGATTTCAAGATTCATCTTCTGTAACAGTTATGATGAGAG 2047
 Db 2246 TTGCTCAGTCTGAGAGAGAGATTTCAAGATTCATCTTCTGTAACAGTTATGATGAGAG 2305
 Qy 2048 TTACAGACAGAGATGATGATGATTTTATGATGAGAGAGAGCTGTGAGAGTGTAGTG 2107
 Db 2306 TTACAGACAGAGATGATGATGATTTTATGATGAGAGAGAGCTGTGAGAGTGTAGTG 2365
 Qy 2108 TTTTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2167
 Db 2366 TTTTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2425
 Qy 2168 AGGCTTGAAGCTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 2227
 Db 2426 AGGCTTGAAGCTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 2485

Qy 2228 CTGCTCTTGACAG 2287
 Db 2486 CTGCTCTTGACAG 2545
 Qy 2288 AGCCAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
 Db 2546 AGCCAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2586
 RESULT 3
 AY410585 2589 bp DNA linear GSS 12-DEC-2003
 LOCUS Mus musculus HCM3933 gene, VIRUAL TRANSCRIPT, partial sequence.
 DEFINITION
 ACCESSION AY410585
 VERSION AY410585.1 GI:39766553
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2589)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferritera,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J., Adams,M.D. and Cargill,M.
 TITLE Interfering nonneutral evolution from human-chimp-mouse orthologous gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PubMed 14671302
 REFERENCE 2 (bases 1 to 2589)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferritera,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J., Adams,M.D. and Cargill,M.
 TITLE Direct Substitution
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment
 FEATURES
 source
 1..2589
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>2589
 /locus_tag="HCM3933"
 ORIGIN
 gene
 Query Match 72.4%; Score 1684.6; DB 29; Length 2589;
 Best Local Similarity 82.8%; Pred. No. 0;
 Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;
 Qy 6 CTTTGTGTGACACAG 65
 Db 267 CTTTGTGTGACACAG 326
 Qy 66 AGGCAATTAATGTAAGTGTGTAAGCACTTCCGCTGACATCCGCTGCCAGTGGGC 125
 Db 327 TGGCATCATTTGTAAGTGTGTAAGCACTTCCGCTGACATCCGCTGCCAGTGGGC 386
 Qy 126 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185
 Db 387 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 446
 Qy 186 TCTTCTTTTCAACAGAGAGATTAATGTAAGTGTGATGATGATGATGATGATGATGATGAT 245
 Db 447 ACTGCTTTTCAACAGAGAGATTAATGTAAGTGTGATGATGATGATGATGATGATGATGAT 506
 Qy 246 ATTATTAATTAAG 305
 Db 507 GTTTTAACTTAAG 566

306 AGAGATGTGAGATTAACGATCACTTTTAAACAATGAATCTTCCACTATCAACCACTTG 365
567 GGAACAGTGAAGATCAACCATCAACCTGACCAAGAGTGGCGCCACCTGCGCCACTG 626
366 TTGGAGTTCTTAATAATTAATTTTCAAGAGCTTTTGAATACTAATTTTGGCAACAAT 425
627 CCGTGAAGTTCTTAATAATCAATCTTCAAGAGCTTTTGAATACTAATTTTGGCAACAAT 686
426 TGAAGCAATTAATTAACCAAAATGATGATGATGATGATGATGATGATGATGATGAT 485
687 TGAAGCAATTAATTAACCAAAATGATGATGATGATGATGATGATGATGATGATGAT 746
486 TTGGCTGAGCTTCACTACTTCCATCTTCAATGATGATGATGATGATGATGATGATGAT 545
747 CTGGCCCGGCTTCAACCACTTCCATCTTCAATGATGATGATGATGATGATGATGATGAT 806
546 CGTTAGCAATTAAGCTTGAAGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 605
807 CGTCAAGCAATTAAGCTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 866
806 TCAGACAGAGAGCAATTAATTTCAAGAACAACTTATGATGATGATGATGATGATGATGAT 665
867 GCAAGACAGAGAGCAATTAATTTCAAGAACAACTTATGATGATGATGATGATGATGATGAT 926
666 TCTTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 725
927 TCTTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 986
726 CAAGACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTAATTAATTAATTAATTA 785
987 AAGAGCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTAATTAATTAATTAATTA 1046
786 GCAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 845
1047 GCAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1106
846 AAGAGCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTAATTAATTAATTAATTA 905
1107 GAG 1166
906 TCTTACAGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 965
1167 TCTTACAGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1226
966 TCTTACAGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1025
1227 GCAACAG 1286
1026 TCAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1085
1287 CCAACAAG 1346
1086 CTTTACCTTCTTCTTCAAG 1145
1347 CTTTACCTTCTTCTTCAAG 1406
1146 ATTGATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1205
1407 GTTGTATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1466
1206 TATGTTTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1265
1467 CAGCGTGAAG 1526
1266 CAATTGATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTA 1325
1527 CAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTGAGAGTCAAGTCTTGAATTAATTA 1586
1326 AGCAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTTGAAGTCTTGAAGTCTTGAAG 1385
1587 GGAATTAACAACAGAGATCAACCACTTGAAGAGCGAGCTTGAAGTCTTGAAGTCTTGAAG 1646

1386 GGTCAACAGAGACACCCAGATAGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1445
1647 GGTGAGTGAAGACACCCAGATAGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1706
1446 CGATGCTATTAATAAATACTGTGTACAGATGAGCTTACCCCAAGTCAAGTGTGTGTGTGTGTGT 1505
1707 TGAATGCAATCAAGAGATCACTGTGTACAGATGAGCTTACCCCAAGTCAAGTGTGTGTGTGTGT 1766
1506 CCGAATCTTAAGGCAACAGCAACTGTCAATGAGCTTGTCAATTAAGATGAGCTTGTCAAGT 1565
1767 CCGAATCTTAAGGCAACAGCAACTGTCAATGAGCTTGTCAATTAAGATGAGCTTGTCAAGT 1826
1566 GAACTGCAAGATGAG 1625
1827 GAACTGCAAGATGAG 1886
1626 CGTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1685
1887 CGTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1946
1686 TGGCAATCAATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1745
1947 CGCAGATCAATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2006
1746 ACAAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1805
2007 GCAAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2066
1806 CTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1865
2067 CTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2126
1866 GCTGAATAACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1925
2127 GCTGAATAACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2186
1926 TAAAGGTTACCACTTGAAG 1985
2187 GAAAGGTTACCACTTGAAG 2246
1986 TTTTGTGAT 2045
2247 TTTTGTGAT 2105
2046 GATTACCAACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2165
2207 AGTACCAACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2225
2106 TGTGTTTCCCAACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2286
2267 TGTGTTTCCCAACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2346
2166 ACAAGGCTTGAAG 2225
2427 CCAAGGCTTGAAG 2286
2226 TCTGCTTCT 2285
2487 CCGTGAACCTTGAAG 2346
2286 AGAGCCAAATCTGTCACTGTCAACAG 2328
2547 AGAGCCAAATCTGTCACTGTCAACAG 2589

RESULT 4
BF66315 874 bp mRNA linear EST 21-DEC-2000
LOCUS 602119233F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276136 5',
DEFINITION mRNA sequence.
ACCESSION BF66315
VERSION BF66315.1 GI:11940210
KEYWORDS EST.


```

Db      2  CAGGTGGAAGAAACCTTGTATACATCCACATTTGCAAGTTGGTCAAAAAGAAACAGAGA 61
QY      1192  GGTGACCATTAATTAATGTTTAAAGCCACTAGATACTGGCTGTGATCTATACCGAGAGA 1251
Db      62  GGTGACCATTAATTAATGTTTAAAGCCACTAGATACTGGCTGTGATCTATACCGAGAGA 121
QY      1252  AATTATGAAGCCGCAATTCATGATACAAATCTATTATTAAGTTACACCGAGCCATGGGC 1311
Db      122  AATTATGAAGCCGCAATTCATGATACAAATCTATTATTAAGTTACACCGAGCCATGGGC 181
QY      1312  ATGCAATAGAGAAAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1371
Db      182  ATGCAATAGAGAAAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
QY      1372  GTCTTACAGCAAAAGGTCAACAGAGACACCCAGATGATGTTGCTGCTGTTGCTGATAT 1431
Db      242  GTCTTACAGCAAAAGGTCAACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 301
QY      1432  CGGAGAGCAAAATAGATGCTATTAATAAATACCTGCTGATGATGATGATGATGATGATGAT 1491
Db      302  CGAAGAGCAAAATAGATGCTATTAATAAATACCTGCTGATGATGATGATGATGATGATGAT 361
QY      1492  CAGTGTGTGTGTGCCCCGCAACCTTAGGCAAAACAGCAAACTGTCATGCGCATTTGCTACA 1551
Db      362  CAGTGTGTGTGTGCCCCGCAACCTTAGGCAAAACAGCAAACTGTCATGCGCATTTGCTACA 421
QY      1552  ATTGGCCCTACAGATTAATCTGCAAGATGGAGAGAGAGCTCTGAGAGGTGGAATCCCTCG 1611
Db      422  ATTGGCCCTACAGATTAATCTGCAAGATGGAGAGAGAGCTCTGAGAGGTGGAATCCCTCG 481
QY      1612  AAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1671
Db      482  AAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541
QY      1672  ATCGCAGATTTGTTGTCACGACATCATATGAGGATG 1707
Db      542  ATCGCAGATTTGTTGTCACGACATCATATGAGGATG 577

RESULT 6
LOCUS   BM751055 527 bp mRNA linear EST 04-MAR-2002
DEFINITION K-EST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5',
            mRNA sequence.
ACCESSION BM751055
VERSION   BM751055.1 GI:19080673
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS  Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
          Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
          Kim,Y.S.
TITLE    21C Frontier Korean EST Project 2001
JOURNAL  Unpublished (2002)
COMMENT  Contact: Kim YS
          Genome Research Center
          Korea Research Institute of Bioscience & Biotechnology
          52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
          Tel: +82-42-860-4470
          Fax: +82-42-860-4409
          Email: yongsung@mail.kribb.re.kr
          Plate: 21 row: B column: 08
          High quality sequence stop: 527.
          Location/Qualifiers
            1..527
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="S9SNU601-21-E08"

```

```

/sex="M"
/tissue_type="testes"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/clone_lib="S9SNU601"
/notes="Organ: Stomach; Vector: pME18-FU3; Site 1: XhoI;
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deacapped
with tabacco acid pyrophosphatase (TAP). The deacapped
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII-digested pME18-FU3 vector. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN

Query Match 22.1%; Score 513.4; DB 12; Length 527;
Best Local Similarity 99.6%; Pred. No. 5.2e-135;
Matches 525; Conservatve 0; Mismatches 1; Gaps 1;

QY      930  GCGTAATGATTTTAAAGTATGAAAGACTTAGCCGTTCAACAAGACTACTCCAGAGCA 989
Db      1  GCGTAATGATTTTAAAGTATGAAAGACTTAGCCGTTCAACAAGACTACTCCAGAGCA 60
QY      990  AAGCAGCGTGAAGTGGAGCACTCATTTGATTAATCAATTAATAAAGCATATGTTCAAG 1049
Db      61  AAGCAGCGTGAAGTGGAGCACTCATTTGATTAATCAATTAATAAAGCATATGTTCAAG 120
QY      1050  GAGCTTCAGAGATGGGGTTTGAGCTTTGATTTCCACTACTGCTTCAGAGAGAT 1109
Db      121  GAGCTTCAGAGATGGGGTTTGAGCTTTGATTTCCACTACTGCTTCAGAGAGAT 180
QY      1110  TTTCGAACAGAAAAGATTCCACCAAGTGGAAAACATTGATTAATCAATCCAAATTTGC 1169
Db      181  TTTCGAACAGAAAAGATTCCACCAAGTGGAAAACATTGATTAATCAATCCAAATTTGC 240
QY      1170  AGATTGGTCCAAAGAAACAAGAGTGCACCTTAATTAAGTTAAACCACTAATACTG 1229
Db      241  AGATTGGTCCAAAGAAACAAGAGTGCACCTTAATTAAGTTAAACCACTAATACTG 300
QY      1230  GCTGTGATCTATACCGCAAGAAATTTATGAGAGCCCAATTCAATGATCAAAATCTATT 1289
Db      301  GCTGTGATCTATACCGCAAGAAATTTATGAGAGCCCAATTCAATGATCAAAATCTATT 360
QY      1290  TAAAGTTACACCGCAATGGGCAATGCAATAGAGAAAGCAATATGATTAAGATGATGA 1349
Db      361  TAAAGTTACACCGCAATGGGCAATGCAATAGAGAAAGCAATATGATTAAGATGATGA 420
QY      1350  CAGACTGAAGGCTTACTTAAGAGTCTTACAGCAAAAGTCAACGACAGACCCAGATAGT 1409
Db      421  CAGACTGAAGGCTTACTTAAGAGTCTTACAGCAAAAGTCAACGACAGACCCAGATAGT 480
QY      1410  TGTCTGTCTGTTGTCAAGTATC-GGAAGGACAAATACGATGCTATT 1455
Db      481  TGTCTGTCTGTTGTCAAGTATC-GGAAGGACAAATACGATGCTATT 527

RESULT 7
LOCUS   BF664941 1495 bp mRNA linear EST 21-DEC-2000
DEFINITION 602119191F1 NIH MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5',
            mRNA sequence.
ACCESSION BF664941
VERSION   BF664941.1 GI:11938836
KEYWORDS EST.
SOURCE   Homo sapiens (human)

```


Wed May 12 09:50:21 2004

us-10-043-774b-1.rst

Page 9

OY		1453	ATTAAAAAATACCTGTGTGAAGATTGGCCCAACCCCAAGCAGTGTGTGATGCCGAAC	1512
OY		1450	ATTAAAAAATACCTGTGTGAAGATTGGCCCAACCCCAAGCAGTGTGTGATGCCGAAC	189
OY		1513	TTAGGCAAAACAGCAAACCTGTCACTGGCCATTGCTTAAGAAATTGCCCTTCAGATGACTCG	1572
OY		150	TTAGGCAAAACAGCAAACCTGTCACTGGCCATTGCTTAAGAAATTGCCCTTCAGATGACTCG	249
OY		1573	AAGATGGAGAGAGAGACTCTGAGAGGGTGAGACAATCCCCCTGAAGCTCGTATGATCGTTGAC	1632
Db		250	AAGATGGAGAGAGAGACTCTGAGAGGGTGAGACAATCCCCCTGAAGCTCGTATGATG- TGCC	308
OY		1633	ATCGATTGTATCATGACATGACACT- GGGCGGAGGTCATCGCAGA- TTTGTTGCCA	1690
Db		309	ATCGATTGTATCATGACATGACACTGGGCGGAGGTCATCGCAGATTGTTGTTGCCA	368
OY		1691	GCATCAATGAAGGGATGAGACCCGCTGTTCTCACGCTGCATATTTCAGATAGAGACAG	1750
Db		369	GCATCAATGAAGGGATGAGACCCGCT- GTTCTCACGCTGCATATTTCAGATAGAGACAG	427
OY		1751	AGCTGTGATAGGGGTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCT- TGAATATGCT- -	1807
Db		428	AGCTGTGATAGGGGTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCTGTGAATAAGCTTG	487
OY		1808	GCAATGATGATCATGCCCAGCCGAGT- CATGCTGACCCGATGCGCGTAGAGACAGCGCAG	1866
Db		488	CAATTGATGATCATGCCCAGCGGAGTCCATCGTACCGAGATGGGTACGAGACGCGCAG	547
OY		1867	CTG- AAAACACTGTGTGAATACGAAGTGGCACAGTTTTGGATTGTCTAAATGCG- ATTG	1924
Db		548	TGAAAAAACCACTGTGAATACGAAGTGGCACAGTTTTGGATTGTCTAAATGCGATTG	607
OY		1925	GTAGAGTTACACACCCTGACTCAACGCTAAATTGTGTGAAGAAAAGATGGAACACCAAGT	1984
Db		608	GTAGAGTTACACACCCTGACTCAACGCTAAATTGTGTGAAGAAAAGATGGAACACCAAGT	667
OY		1985	TT 1986	
Db		668	TT 669	
RESULT 9				
BF080923				
LOCUS	BF080923	533 bp	mRNA	linear EST 18-OCT-2000
DEFINITION	233690 MAREC 2PBG Sus scrofa cDNA 5', mRNA sequence.			
ACCESSION	BF080923			
VERSION	BF080923.1			
KEYWORDS	EST.			
SOURCE	Sus scrofa (pig)			
ORGANISM	Sus scrofa			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eubacteria; Cetartiodactyla; Suina; suidae; Sus.			
	1 (bases 1 to 533)			
	Fahnenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,			
	Vallier,J., Wise,T., Rohrer,G.A., Petrea,G., Sultana,R.,			
	Quackenbush,J., and Keefe,J.W.			
	Porcine gene discovery by normalized cDNA-library sequencing and			
	EST cluster assembly			
	Mamm. Genome 13 (8), 475-478 (2002)			
TITLE				
JOURNAL				
MEDLINE				
PubMed				
COMMENT				
	Contact: Smith TPB			
	USDA, ARS, US Meat Animal Research Center			
	PO. Box 166, Clay Center, NE 68933-0166, USA			
	Tel.: 402 762 4366			
	Fax: 402 762 4390			
	Email: smith@email.marc.usda.gov			
	Single pass sequencing. Bases called and alt trimmed with phred			
	v0.980904.e. Vector identified by cross_match with the -mismatch 18			
	and -mismatch 12 options.			
	PCR Primers			
	FORWARD: AGAAAAAGCTATGACAT			
	BACKWARD: GTTTCCAGTCAAGAC			

FEATURES	Source	Location/Qualifiers
Plate: 53	row: 0	column: 21
Seq primer:	ATTTAGGAGACACTATAG.	
1..533		
/organism="Sus scrofa"		
/mol_type="mRNA"		
/db_xref="taxon:9823"		
/citation_type="pooled"		
/lab_host="DH10B"		
/clone_id="MARC_2P1G"		
/note="Vector: pCMV Sport6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."		
ORIGIN		
Query Match	20.2%; Score 471.2; DB 10; Length 533;	
Beet Local Similarity	92.9%; Pred. No. 6.1e-123;	
Matches	494; Conservative 0; Mismatches 38; Indels 0; Gaps 0;	
QY	1132 CAAGTGGAAAACATTTGATTTGATTCACCAATTGGAGATGGTCCAAAGAAACAGA	1199
DB	2 CAAGGTGAAAACATGGGATTTACATCCAAATTTGACAGATGGTCCAAAGAAACAGA	61
QY	1192 GGTGCACCATTAATTAAGTGTAAAGCCACTAGATTAAGTGGCTGTGATCTATACGGAGA	1251
DB	62 GGTGCACCATTAATTAAGTGTAAAGCCACTAGATTAAGTGGCTGTGATCTATACGGAGA	121
QY	1252 AATTATAGAGGCAATTCATTTGATTCACCAATCTATTTAAAGTTACACGAGCATGGGC	1311
DB	122 AATTATAGAGGCAATTCATTTGATTCACCAATCTATTTAAAGTTACACGAGCATGGGC	181
QY	1312 ATGCCAAATGAAAAAGCAATATATGATTGAATGGATGACAGAACTGAAAGCTACTTAAAG	1371
DB	182 ATACAAATGAAAAAGCAATATATGATTGAATGGATGATTAAGTGAAGCTACTTAAAG	241
QY	1372 GTCTTACAGAAAGGTTCACAGCAGACACCAGATGTTGTCGTCTGTTTCAAGTAT	1433
DB	242 GTCTTACAGCAAAAGTTACATCGGATTAACCCAGATGTTGTTGTCGTCTGTTTCAAGTAT	301
QY	1432 CGGAAGACAAATTCAGATGCTATTTAAAAAATACCTGTGTACAGATTGGCCCTACCCCAAGT	1491
DB	302 CGAAAGACAAATTCAGATGCTATTTAAAAAATACCTGTGTACAGATTGGCCCTACCCCAAGT	361
QY	1492 CAGTGTGTGGTGGCCGCAACCTTATAGGCAACAGAACTGTCAAGCCATTGCTACAAAG	1551
DB	362 CAGTGTGTGGTGGCCGCAACCTTATAGGCAACAGAACTGTCAAGCCATTGCTACAAAG	421
QY	1552 ATTGCCCTTACAGATGAATCTGCAGATGGAGAGAGAGCTCTGGAGAGGTGGACATCCCTCTG	1611
DB	422 ATCCGCTTACAGATGAATCTGCAGATGGAGAGAGAGCTCTGGAGAGGTGGACATCCCTCTG	481
QY	1612 AAGTCGTGATGATCGTTGGCATGATTTGTTACATGACATGACAGAGCTGGGC	1663
DB	482 AAGTCGTGATGATCGTTGGCATGATTTGTTACATGATACCAAGCTGGGC	533
RESULT 10		
LOCUS	AKO30116	3555 bp mRNA linear HTC 18-SEP-2003
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone4932443D15 product:plwi like homolog 1 (Drosophila)-like, full insert sequence.	
ACCESSION	AKO30116	
VERSION	AKO30116.1	GI:26326094
KEYWORDS	HTC; CAP trapper.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 Carninci, P. and Hayashizaki, Y.	
AUTHORS	High-efficiency full-length cDNA cloning	
TITLE	Meth. Enzymol. 303, 19-44 (1999)	
JOURNAL		

MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komo, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

QY 724 CCAGAGGACCTTTAAGAAAGCCGAGCTTGAAGTCAAGCTTCTTAAGAAATCTACAGG 783
 Db 1447 CCTAAAGACAGCTTGTGATGTCGAGAGGGAAGAAATCACTTCTGGAATCTACAGC 1506
 QY 784 AAGGAATACAAACGAAGATCAACGACTTGAAGAGCTGCTTGTGACAGCCCA-- 841
 Db 1507 AAAAATATGGGATCAAGTCAAGAAAGATGACAGCCGCTGCTGATTCACCGCCAGT 1566
 QY 842 -AGAGAGCGGGGCGCTGGGGGACACTGCAAGGCGCTGCTGATTCCTGAGCTC 900
 Db 1567 GAGAGACAGAAATACATGGCATGTGTGTAAGGGCGAGATCCGTGCTGCGCCAGCTC 1626
 QY 901 TGTATCTTACAGCTCTACTGATTAATGGTAAATTTTAAGTGAAGAAAGCTTA 960
 Db 1627 TCTTCATGACGGGGATCCCTGAGAAAGTGAAGAGCTTCAAGGCGCATGAAGAGCTTG 1686
 QY 961 GCGGTTCAACAGACTAATCCAGAGCAAGGAGCGGTGAAGTGGAGCACTCATTTGAT 1020
 Db 1687 ACTGACAGATTTAACTGAGCCCAAGAGACACACAGCTGCTTGAATGCTGCTGAG 1746
 QY 1021 TACATTCATAAAGATTAATGTTCAAGAGAGCTTGAAGCTGGGCTTGAAGCTTGAAT 1080
 Db 1747 AGAATTTCAAAAACGAGGAGCGAGCATGAGCTGACCGCTGGGGCTCAAGTCTCAT 1806
 QY 1081 TCCAACTTACTGCTCTCTGAGAAAGATTTTGCAACAGAAAGATTCACCAAGTGA 1140
 Db 1807 AAAGATGTCACAAAGATTTGAAGTGGCTTCTGCAATGAGAGATCAATTAAGAAC 1866
 QY 1141 AAAACATTTGATTAACAATCCAAATTTGCAATTTGCTCAAAAAGAAACAGAGTGAACA 1200
 Db 1867 ACTTCATTTG--TCACATCGGAGGAGCTGAACTGGGTTAAGAAAGTGAACAGAGATGCT 1923
 QY 1201 TTAATTAATGTTAAGCCACTGATTAAGTGGCTGATCTATAGCCGAAGAAATTAAGAA 1260
 Db 1924 TCCATTTCACTATTTCCATGCTATTTCTGGGACCTCTTTTATCCAAAGAGCAATGAGC 1983
 QY 1261 GCAGCCAAATTCATGATCAAAATCTATTATTAAGTTACACAGCCATGGGAGCAATG 1320
 Db 1984 CAAGCCAGAGAACTGTTAACTGTTGAAGAAAGATTCGCGGCCCATTTGGCATCGCAT 2043
 QY 1321 AGAAAGCAATATATTTGAAGT--GATGACAGAACTGAAGCCTATTGAAGTCTTA 1377
 Db 2044 AGCCCCCAGCGCTGGTTGAGCTGAAGATGACCAAAAGAGACCTATATACAGACCTT 2103
 QY 1378 CAG-----CAAAAGTCAACAGCAGACACCCAGATGTTGCTGCTGTTGCAAGTAT 1431
 Db 2104 CAGTCTTAATCTGAGAGTTGAGGAGAAAGTAAAGATGCTGTTGATCATATGAGGACA 2163
 QY 1432 CGGAAGCAAAATACATGATCTATTAATAAATCTGTGTACAGATGCTGCTTACCCCAAGT 1491
 Db 2164 CGTGTATGATCTCTATGAGCCATCAAGAGCTGTGCTGCTGCACTCCCGAGTCTCA 2223
 QY 1492 CAGTGTGAGTGGCCGAACTTAGGCAAAACAGAACTGTCAATGGCCATTGCTACAAAG 1551
 Db 2224 CAGGTCAATATGTCGAAACCTTGGTCAAGCCCAACAGGCTTGGAGGTGGCTAGAAA 2283
 QY 1552 ATTGCCCTACAGATGAATGCAAGATGGAGAGAGCTTGGAGGTGGACATCCCTCTG 1611
 Db 2284 ATTTTACTTCAGATGAATGTAACCTGGGTGGTGGCTGCTGGGAGTGAATATCCGCTG 2343
 QY 1612 AAGCTGATGATGCTGGGCAATGATTTTACATGATGACAGCTGGGGCGAGTCA 1671
 Db 2344 AAACAACATATATGATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 2403
 QY 1672 ATGCGAGATTTGTTGCCAGCATCAATGAAGGATGACCGGTGTTCTCAAGCTGACATA 1731
 Db 2404 GTGGTGGCTTGGTGGCCAGCATAAATTCACATTCACCAATGTTACTGAGGGTGGTG 2463
 QY 1732 TTTTCAGATGAGAGAGAGAGCTGGTAATGGGCTCAAAAGTGCCTGCAAGCGGCTG 1791
 Db 2464 TTCAGATGCCACATAGAGAGATGTGAGACAGCTGAAAGCTTGGCTGTTGGGTTCTTG 2523

QY 1792 AGGGCTTGGAATAGTGCATGATGATACATGCCCCAGCCGATCATGCTGATCCGATGAGC 1851
 Db 2524 AAAAATATTTATGAGTGAACATGCTGCTCCAGAGAAATATGCTGATACGAGATGGA 2583
 QY 1852 GTAGAGACGGCCAGCTGAAAACACTGTGAATCTACAGATGCCACAGTTTGGATGCT 1911
 Db 2584 GTGTCTGATGCGCAGCTAAAGACAGTTGCAACTACAGATCCCTGAGCTGCAAGATGT 2643
 QY 1912 CTAAATCATTTGTAGAGGTATACACCTGATCAAGGTAAATGTTGGTGAAGAAAAGA 1971
 Db 2644 TTGAAGCTTTGATAC--TACACCCCAAGATGATGATGTTGTGATGTAAGAAAGA 2700
 QY 1972 GTGAACACCAATTTTGTGCTGATGCTGAGAGAGACTTCAGAAATCCACTTCTGAGACA 2031
 Db 2701 ATCAGACCAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 QY 2032 GTTATGATGATGAGGTATCCAGACCAAGATGATGATTTTATAGTGAAGCCAGCT 2091
 Db 2761 GTGTGATGATATACCATTAACAGCTGTGATGATGATGATTTCTACTTCTTGGCCATAT 2820
 QY 2092 GTGAGAGATGATGATGATTTCTCCACACATTAATGCTATGATGATGATGATGATGAT 2151
 Db 2821 GTGCGACAGGGCTGTGAGCATGATCTACACTGATGATGATGATGATGATGATGATGAT 2880
 QY 2152 AAGCCAGACCAATACAGCGCTTGAATCAAGCTGTGCAATGCTATTAACATGAGCA 2211
 Db 2881 AGCCCTGATCACTGAGAGAGTGTGACTTCAACATATGCAATGATGATGATGATGATGAT 2940
 QY 2212 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2271
 Db 2941 GGTACCATCCAGATTTCAAGCTCTTGAAGATGATGATGATGATGATGATGATGATGAT 3000
 QY 2272 CAGAGATTTACAGAGAGCCAAATCTGATGATGATGATGATGATGATGATGATGATGAT 2328
 Db 3001 CAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3057

RESULT 11
 AA969938
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 462)
 NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 tumor gene index
 Unpublished (1997)
 Contact: Robert Straube, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 This clone is available royalty-free through JLNH; contact the
 IMAGE Consortium (infoimage.jlnh.gov) for further information.
 Insert length: 521 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham.
 Location/Qualifiers
 1. 462
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1581212"
 /lab_host="DH10B"
 /clone_id="Soares_NFL_T_GBC_S1"
 /note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with
 a modified polylinker; Site: 1; Not 1; Site 2; Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung Ndh19w, testis NHT, and B-cell
 NCI-GAP_GCB1) were mixed, and ss circles were made in

FEATURES

source

ORIGIN

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

Query Match 19.0%; Score 441.4; DB 9; Length 462;
Best Local Similarity 99.6%; Pred. No. 1.9e-114;
Matches 453; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 1714 TGGTTCACGCTGCAATTTTCAGATAGAGACAGAGCTGTAGATGGCTCAAGTC 1773
1 TGGTTCACGCTGCAATTTTCAGATAGAGACAGAGCTGTAGATGGCTCAAGTC 60
Qy 1774 TGCCTGACAGCGGCTTGAAGGCTTGAATAGCTGACATGATACATGCCAGCCGATC 1833
61 TGCCTGACAGCGGCTTGAAGGCTTGAATAGCTGACATGATACATGCCAGCCGATC 120
Qy 1834 ATGCTGACAGCGGCTTGAAGGCTTGAATAGCTGACATGATACATGCCAGCCGATC 1893
121 ATGCTGACAGCGGCTTGAAGGCTTGAATAGCTGACATGATACATGCCAGCCGATC 179
Qy 1894 CCACAGTTTGTGATGTTAAATCCATGTTAGAGGTTACAACTTACATGAGGTA 1953
180 CCACAGTTTGTGATGTTAAATCCATGTTAGAGGTTACAACTTACATGAGGTA 239
Db 1954 ATTGTGTGAG 2013
240 ATTGTGTGAG 299
Qy 2014 AATCCACTTCTGAGACAGTTATGATGATGATGATGATGATGATGATGATGATG 2073
300 AATCCACTTCTGAGACAGTTATGATGATGATGATGATGATGATGATGATGATG 359
Db 2074 TTTATCGTGAAGCAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2133
360 TTTATCGTGAAGCAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Qy 2134 TATGACAAAGCGGCTGAGAGCAGACACATACA 2168
420 TATGACAAAGCGGCTGAGAGCAGACACATACA 454
Db

RESULT 12 818 bp mRNA linear EST 15-MAY-2001
Bg773137
LOCUS 602721862P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838918 5',
DEFINITION mRNA sequence.

ACCESSION Bg773137 GI:14083790
VERSION Bg773137
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS (bases 1 to 818)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgsb@remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMU at:
http://image.llnl.gov
Plate: LLM10773 row: k column: 15

FEATURES
source
High quality sequence stop: 789.
Location/Qualifiers
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4838918"
/lab_host="DH10B"
/clone_id="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gagag); Oligo-dT primed using primer 5'-TTTTTTTCTTTTCTT-3', size selected for average insert size 2.2 kb and normalized for RGT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGR, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 18.8%; Score 438.8; DB 12; Length 818;
Best Local Similarity 97.3%; Pred. No. 1.4e-113;
Matches 500; Conservative 0; Mismatches 7; Indels 7; Gaps 5;

Qy 8 TTGTGTGAACAAGGAG 67
Db 305 TTGTGTGAACAAGGAG 364
Qy 68 GCATTAATGTAAGGTTAG 127
Db 365 GCATTAATGTAAGGTTAG 424
Qy 128 TATATCAATCACTTACATTAACCACTGATGAGAGAGAGAGAGAGAGAGAGAGAG 187
Db 425 TATATCAATCACTTACATTAACCACTGATGAGAGAGAGAGAGAGAGAGAGAGAG 484
Qy 188 TTTCTTTTCAACAG 247
Db 485 TTTCTTTTCAACAG 544
Qy 248 TTTTACTTAAAG 307
Db 545 TTTTACTTAAAG 604
Qy 308 AGAGATGAG 366
Db 605 AGAGATGAG 664
Qy 367 TTGCAATTTCTATATATATATTTTCAAGAGAGC-TTTTGAATATCATGAATTTC--AAC 422
Db 665 TTGCAATTTCTATATATATATTTTCAAGAGAGC-TTTTGAATATCATGAATTTC--AAC 724
Qy 423 AATGAG 481
Db 725 AATGAG 784
Qy 482 TGAATTTGG-CCTGAGCTTCACTTCAATCCCTTC 514
Db 785 TGAATTTGGCCTTGCTTCACTTCAATCCCTTC 818

RESULT 13 642 bp mRNA linear EST 16-OCT-2002
BUBS4142
LOCUS 10402932 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6621314
DEFINITION 5', mRNA sequence.

ACCESSION BUBS4142 GI:24039108
VERSION BUBS4142
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

QY 1797 TTGAAATAGCTGCAATAGTACATGCGCCGAGATCATGTGTAACCGGATGCGGTAGG 1856
 DB 121 TTGAGAGTGGCTGCATGAATGATACATGCGCCGAGATCATGTGTAACCGGATGCGGTAGG 180
 QY 1857 AGACGGCCGAGCTTAAACACTGTGTGAACATGAGAGGCAAGTTTGGATTGCTAAA 1916
 DB 181 GAGCGGGAGCTTAAGACCTGTGTCAATTATGAGGCCACAGTTCCTGATTTGCTCAA 240
 QY 1917 ATCCATGTGTAGAGGTACAAACCTGACTAGCTAGCTGATTTGGTGAAGAAAAGTGA 1976
 DB 241 GTGAGTGGGAGAGGTACAAACCAAGACTGACTGTAATGTGTGAAGAAAGGCTGCA 300
 QY 1977 CACCAATTTTTCCTGATGCTGAGAGGAAGCTTCAAGAAATCCACTTCCCTGAAAGCTAT 2036
 DB 301 TGCCAGGTTTTTGTGCTAGTGTGGGAGAGCTTCAAGAACTTCCAGGAGAGTCAAT 360
 QY 2037 TGAATGAGAGTTTACCAAGACCAAGATGTATGACTTTTATGTGAGCCAGGCTGTAG 2096
 DB 361 CGATGTGGAAGTACCAAGACCAAGAGTGTATGACTTTGTCACTGTGAGCCAGGCTGTAG 420
 QY 2097 AAGGTGAGTGTGTCTCCACACTTACATGTCTATGATGAGCAAGCGGCTGAGGCC 2156
 DB 421 AAGGGAGAGTGTGCCCCAACACACTCAATGTCTATGATGAGCAAGCGGCTGAGGCC 480
 QY 2157 AGACCAATACAGGCGCTTGAACCTTCAAGCTGTGCAATCTATTATTAACAAGTGGCAGGTGT 2216
 DB 481 CGACCAATTCAGGCGCTGATACATGCAAGCTGTGTCAAGTGTATTAATTGGCTGAGGT 540
 QY 2217 CATTGCTGTCTGCTGCTCTTGGCA 2240
 DB 541 CATGCGAGTTCCTGCACTTGGCA 564

RESULT 15
 AL041473/c 413 bp mRNA linear EST 04-SEP-2003
 LOCUS DKFZP434E2417.81 434 (synonym: htes3) Homo sapiens cDNA clone
 DEFINITION DKFZP434E2417.3, mRNA sequence.
 AL041473
 AL041473.1 GI:5420824
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 413)
 Bloecher, H., Boecher, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
 Wiemann, S.
 EST (Bioecker, et al.)
 TITLE Unpublished (1999)
 JOURNAL Contact: MIPS
 COMMENT MIPS
 Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 3' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de, Ltd.,
 Braunschweig/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 r1 sequence also available.
 This clone (DKFZP434E2417) is available at the RZPD in Berlin.
 Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1..413
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZP434E2417"
 /feature_type="est"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="434 (synonym: htes3)"

ORIGIN /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
 Query Match 17.7%; Score 413; DB 9; Length 413;
 Best Local Similarity 100.0%; Pred. No. 2.3e-106;
 Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1637 ATTGTTACATGATGATGACGCTGGGGAGGTCAATCGAGAGATTGTTGCCAGCATCA 1696
 DB 413 ATTGTTACATGATGATGACGCTGGGGAGGTCAATCGAGAGATTGTTGCCAGCATCA 354
 QY 1697 ATGAAGGATGACCCGCTGTTCTACGCTGECATATTTCAAGATGAGACAGAGAGCTGG 1756
 DB 353 ATGAAGGATGACCCGCTGTTCTACGCTGECATATTTCAAGATGAGACAGAGAGCTGG 294
 QY 1757 TAGATGGCTCAAGTCTGCTGCAAGGGCTGAGGGCTTGAATATGCTCAATAGT 1816
 DB 293 TAGATGGCTCAAGTCTGCTGCAAGGGCTTGAATATGCTCAATAGT 234
 QY 1817 ACATGCCAGCGGATCATGCTGTACCGCGATGGCGTATGAGAGACGGCCAGCTGAAAACAC 1876
 DB 233 ACATGCCAGCGGATCATGCTGTACCGCGATGGCGTATGAGAGACGGCCAGCTGAAAACAC 174
 QY 1877 TGGTGAATACGAAGTGGCAAGTTTGGATTTGCTTAAATTCATTGTGTAGAGTTACA 1936
 DB 173 TGGTGAATACGAAGTGGCAAGTTTGGATTTGCTTAAATTCATTGTGTAGAGTTACA 114
 QY 1937 ACCCTAGACTAAGCTAATGTGTGTAAGAAAGATGAAACACAGATTTTTCCTCAGT 1996
 DB 113 ACCCTAGACTAAGCTAATGTGTGTAAGAAAGATGAAACACAGATTTTTCCTCAGT 54
 QY 1997 CTGAGGAAGACTTCAGATTCATCTCTGCAACAAGTTATGATGATAGAGTT 2049
 DB 53 CTGAGGAAGACTTCAGATTCATCTCTGCAACAAGTTATGATGATAGAGTT 1

Search completed: May 8, 2004, 15:48:52
 Job time : 5991 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:06:00 ; Search time 60 seconds
(without alignment)
3649.572 Million cell updates/sec

Title: US-10-043-774B-2
Perfect score: 4102
Sequence: 1 MIFGNTNRQNDHYKESKMG.....VGQSHREPNLSLNRLYYL 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: Geneseqp19808:*\n2: Geneseqp19808:*\n3: Geneseqp2000s:*\n4: Geneseqp2001s:*\n5: Geneseqp2002s:*\n6: Geneseqp2003as:*\n7: Geneseqp2003bs:*\n8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4102	100.0	775	ADB94987	ADB94987 Human hae
2	4061	99.0	861	AAV90235	AAV90235 Human p1w
3	3957	96.5	862	AAV90234	AAV90234 Mouse p1w
4	3883	94.7	829	ADB65285	ADB65285 Human pro
5	2764	67.4	523	ABP67995	ABP67995 Human col
6	2764	67.4	523	ABU61933	ABU61933 Human Hiw
7	2101.5	51.2	852	ADB65492	ADB65492 Human pro
8	1468	35.8	866	ABU61932	ABU61932 Drosophila
9	1468	35.8	866	ABU61933	ABU61933 Drosophila
10	1434	35.0	844	ADB94987	ADB94987 Fruitfly
11	1434	35.0	844	ADB94988	ADB94988 Fruitfly
12	1428.5	34.8	843	ABU61933	ABU61933 Drosophila
13	1416.5	34.5	843	AAV90233	AAV90233 Drosophila
14	1389.5	33.9	580	AAU07866	AAU07866 Polypepti
15	1324.5	33.3	498	ABP43865	ABP43865 Human mRN
16	1324.5	33.3	498	ABU61938	ABU61938 Human ful
17	1295.5	31.6	824	ABU61935	ABU61935 C. elegan
18	1230	30.0	530	ABU61935	ABU61935 C. elegan
19	1162	28.3	722	ABU61935	ABU61935 C. elegan
20	962	22.5	371	AAV92722	AAV92722 Human pro
21	802	15.6	642	ADA54973	ADA54973 Human pro
22	652	15.9	130	AAU58977	AAU58977 Breast an
23	580	14.1	341	ABG05957	ABG05957 Novel hum
24	515.5	12.6	915	ADB17496	ADB17496 Soybean p
25	510	12.4	868	AAU42119	AAU42119 Arabidops

26	510	12.4	1048	3	AAU42118	AAU42118 Arabidops
27	510	12.4	1048	6	ADB17512	ADB17512 A. thalia
28	508.5	12.4	859	5	ABP64718	ABP64718 Human pro
29	508.5	12.4	860	5	ABG97473	ABG97473 Human NAA
30	507.5	12.3	1101	6	ADB17510	ADB17510 Rice post
31	506.5	12.3	857	7	ADB81572	ADB81572 Human euk
32	506	12.3	186	7	ABU61932	ABU61932 Human par
33	503	12.3	906	3	AAU52594	AAU52594 Arabidops
34	503	12.3	906	3	AAU51563	AAU51563 Arabidops
35	503	12.3	983	3	AAU51562	AAU51562 Arabidops
36	503	12.3	988	3	AAU51561	AAU51561 Arabidops
37	503	12.3	1064	3	AAU52593	AAU52593 Arabidops
38	500	12.2	950	4	ABU67046	ABU67046 Drosophila
39	500	12.2	984	4	ABU62573	ABU62573 Drosophila
40	500	12.2	984	4	ABU67045	ABU67045 Drosophila
41	500	12.2	984	5	ABU70016	ABU70016 Larval vi
42	498.5	12.2	1058	6	ADB17494	ADB17494 Soybean p
43	497	12.1	816	6	AAU57044	AAU57044 Human nuc
44	497	12.1	860	4	AAU60292	AAU60292 Human pro
45	497	12.1	860	4	AAU94291	AAU94291 Human pro

ALIGNMENTS

RESULT 1
ADB94987
ID ADB94987 standard; protein; 775 AA.

AC ADB94987;
DT 04-DEC-2003 (first entry)
DE Human haematopoietic growth regulatory protein, HIWI.
XX Human, haematopoietic growth regulatory protein; HIWI; cytostatic;
XX stem cell therapy; bone marrow; peripheral blood;
XX CD34+ haematopoietic stem cell; leukaemia; cancer.
XX Homo sapiens.
OS US2003129599-A1.
PN 10-JUN-2003.
PD 10-JUN-2003.
PE 10-JAN-2002; 2002US-00043774.
PF 10-JAN-2002; 2002US-00043774.
PR 10-JAN-2002; 2002US-00043774.
PS (SHAR/) SHARMA A.
PS (HOFF/) HOFFMAN R.
XX Sharma A, Hoffman R;
XX WPI; 2003-744346/70.
XX N-PSDB; ADB94986.
XX New hiwi nucleic acids, for screening compounds which modulate hiwi,
XX useful as cancer therapeutic.
XX Claim 2; Fig 1; 39pp; English.
XX The invention relates to a nucleic acid comprising a nucleotide sequence
XX encoding a human hiwi protein (a haematopoietic growth regulatory gene).
XX Also included are the hiwi protein, the hiwi cDNA (used as a probe), a
XX recombinant expression construct comprising the hiwi cDNA, a cell culture
XX transformed with the recombinant expression construct, screening a
XX compound for modulating human hiwi gene activity in cells expressing the
XX human hiwi gene product, a cell membrane preparation (or cytosolic
XX preparation or nuclear preparation) comprising a human hiwi gene product
XX (or derivative) and identifying a compound that induces or increases hiwi
XX gene expression in mammalian cells. Culturing bone marrow or peripheral
XX blood culture in the presence of compounds identified by the methods is

CC useful for increasing retention of primitive CD34+ haematopoietic stem
 CC cells. Alternatively the method comprises introducing a retroviral or
 CC lentiviral expression construct into the cells. The compounds are useful
 CC for haematopoietic stem cell development and maintenance and for
 CC inhibiting leukemia cell growth and treating cancer. The present
 CC sequence represents the hwi protein.

XX
 XX
 SQ Sequence 775 AA;

Query Match 100.0%; Score 4102; DB 7; Length 775;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIFGVTNRQNLVHVKESKTSSGIIYRLSTNHFRLTSRPMALVYHIDNPLMEARRR 60
 DB 1 MIFGVTNRQNLVHVKESKTSSGIIYRLSTNHFRLTSRPMALVYHIDNPLMEARRR 60
 QY 61 SALLFQHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRITITLTNELPPTSP 120
 DB 61 SALLFQHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRITITLTNELPPTSP 120
 QY 121 TCIQFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLC 180
 DB 121 TCIQFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLC 180
 QY 121 TCIQFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLC 180
 DB 121 TCIQFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLC 180
 QY 181 TDVSHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQ 240
 DB 181 TDVSHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQ 240
 QY 181 TDVSHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQ 240
 DB 181 TDVSHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQ 240
 QY 241 NPKSTFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPEL 300
 DB 241 NPKSTFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPEL 300
 QY 241 NPKSTFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPEL 300
 DB 241 NPKSTFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPEL 300
 QY 301 CYLIGLIDPKRNDPNVMKDLAVHTRLTPEQRQREVRGLIDYHKNQVQRELDWGLSFD 360
 DB 301 CYLIGLIDPKRNDPNVMKDLAVHTRLTPEQRQREVRGLIDYHKNQVQRELDWGLSFD 360
 QY 301 CYLIGLIDPKRNDPNVMKDLAVHTRLTPEQRQREVRGLIDYHKNQVQRELDWGLSFD 360
 DB 301 CYLIGLIDPKRNDPNVMKDLAVHTRLTPEQRQREVRGLIDYHKNQVQRELDWGLSFD 360
 QY 361 SNLSFSGRILOTEKIHGGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420
 DB 361 SNLSFSGRILOTEKIHGGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420
 QY 361 SNLSFSGRILOTEKIHGGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420
 DB 361 SNLSFSGRILOTEKIHGGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420
 QY 421 AANSLIOWLFKVTTPAMQKRAKALIEVDRTKAYRLVLOOKTADQIIVCLSSNRKD 480
 DB 421 AANSLIOWLFKVTTPAMQKRAKALIEVDRTKAYRLVLOOKTADQIIVCLSSNRKD 480
 QY 421 AANSLIOWLFKVTTPAMQKRAKALIEVDRTKAYRLVLOOKTADQIIVCLSSNRKD 480
 DB 421 AANSLIOWLFKVTTPAMQKRAKALIEVDRTKAYRLVLOOKTADQIIVCLSSNRKD 480
 QY 481 KYDAIKKYLCTDCPTSPQCVVARTLGGQOTVMAATKIALQMKMGSELNRVDIPUKLV 540
 DB 481 KYDAIKKYLCTDCPTSPQCVVARTLGGQOTVMAATKIALQMKMGSELNRVDIPUKLV 540
 QY 481 KYDAIKKYLCTDCPTSPQCVVARTLGGQOTVMAATKIALQMKMGSELNRVDIPUKLV 540
 DB 481 KYDAIKKYLCTDCPTSPQCVVARTLGGQOTVMAATKIALQMKMGSELNRVDIPUKLV 540
 QY 541 MIVGIDCYHMTAGRRSIAGFVASINEGMTWFSRCLTQDGGQELVNDGLKYLQALRAW 600
 DB 541 MIVGIDCYHMTAGRRSIAGFVASINEGMTWFSRCLTQDGGQELVNDGLKYLQALRAW 600
 QY 541 MIVGIDCYHMTAGRRSIAGFVASINEGMTWFSRCLTQDGGQELVNDGLKYLQALRAW 600
 DB 541 MIVGIDCYHMTAGRRSIAGFVASINEGMTWFSRCLTQDGGQELVNDGLKYLQALRAW 600
 QY 601 NSCNMYMSRIIVVDGVDGQDKTLVYVEVPOFADCLKSIGRGYNPLFIIVVKEKENT 660
 DB 601 NSCNMYMSRIIVVDGVDGQDKTLVYVEVPOFADCLKSIGRGYNPLFIIVVKEKENT 660
 QY 601 NSCNMYMSRIIVVDGVDGQDKTLVYVEVPOFADCLKSIGRGYNPLFIIVVKEKENT 660
 DB 601 NSCNMYMSRIIVVDGVDGQDKTLVYVEVPOFADCLKSIGRGYNPLFIIVVKEKENT 660
 QY 661 REFASGGRLONPFLPGTVIDEVTRPEWYDFEIVSQAVRSQSVSETHNVIVYDNGSLKPD 720
 DB 661 REFASGGRLONPFLPGTVIDEVTRPEWYDFEIVSQAVRSQSVSETHNVIVYDNGSLKPD 720
 QY 661 REFASGGRLONPFLPGTVIDEVTRPEWYDFEIVSQAVRSQSVSETHNVIVYDNGSLKPD 720
 DB 661 REFASGGRLONPFLPGTVIDEVTRPEWYDFEIVSQAVRSQSVSETHNVIVYDNGSLKPD 720
 QY 721 HIGRLTYLCHITVYMPGVIRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775
 DB 721 HIGRLTYLCHITVYMPGVIRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775
 QY 721 HIGRLTYLCHITVYMPGVIRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775
 DB 721 HIGRLTYLCHITVYMPGVIRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775

RESULT 2

AA90235 standard; protein; 861 AA.

AC AA90235;

DT 29-AUG-2000 (first entry)

XX Human piwi protein, designated hwi.
 DE
 XX piwi family protein; piwi; miwi; hwi; gene therapy; tissue dystrophy;
 KW anaemia; immunodeficiency; male infertility; human.
 XX
 XX Homo sapiens.

Key Location/Qualifiers

FT Misc-difference 76 /label= 'Leu, Ile' /note= "encoded by NTA"

FT Misc-difference 303 /label= 'Leu, Ile' /note= "encoded by NTA"

FT Misc-difference 735 /label= 'Leu, Ile' /note= "encoded by NTA"

FT WO20032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US028764.

XX 04-DEC-1998; 98US-0110901P.

XX (UYDU-) UNIT DUKE.

XX Lin H;

XX WPI; 2000-412085/35.

XX N-PSDB; AAA07588.

XX piwi family nucleic acids, polypeptides, and antibodies, useful in gene
 PT therapy of diseases such as cancer and in various research and diagnostic
 PT applications.

XX claim 4; Page 189-194; 201pp; English.

XX This sequence represents the human piwi family protein, designated hwi.
 CC The piwi family nucleic acids and polypeptides are used in gene therapy
 CC of diseases such as cancer and also in various research and diagnostic
 CC applications. The sequences can also be used to treat tissue dystrophy,
 CC anaemia, immunodeficiency, and male infertility

XX Sequence 861 AA;

Query Match 99.0%; Score 4061; DB 3; Length 861;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 768; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GVTNRQNLVHVKESKTSSGIIYRLSTNHFRLTSRPMALVYHIDNPLMEARRRSL 63

DB 90 GVTNRQNLVHVKESKTSSGIIYRLSTNHFRLTSRPMALVYHIDNPLMEARRRSL 149

QY 64 LFOHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRITITLTNELPPTSPCL 123

DB 150 LFOHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRITITLTNELPPTSPCL 209

QY 124 QFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLCYV 183

DB 210 QFYNIIFRRLKIKIMLQOIGRNYNPNPDIIPSHRLVWPGFTTSLQYENSIMLCYV 269

QY 184 SHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQNK 243

DB 270 SHKVLRSFTVLDPMNFYHOTESEKFOVSKEKELGLVLTKNKTYRVDDIDWDQNK 329

QY 244 STFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPELCH 303

DB 330 STFKKADGSEVSFLYYRKYNOEITDLKQPVLSQPRRRPGGTLPGPAMILPELCH 389

QY 304 TGLIDPKRNDPNVMKDLAVHTRLTPEQRQREVRGLIDYHKNQVQRELDWGLSFD 363


```

Db      390 TGLTDMKNDNFVMDLAHVRFLTPBQRORFVGLIDYIHKNDVQRELRDMLSDSNL 449
Qy      364 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 423
Db      450 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 509
Qy      424 SLIQNLFKVTPAMGMQMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 483
Db      510 SLIQNLFKVTPAMGMQMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 569
Qy      484 AIKKYLTCTDPTSCQCVARTLGKQOTVMAITKIALQNNCKMGGELMRVDIPLKLVMI 543
Db      570 AIKKYLTCTDPTSCQCVARTLGKQOTVMAITKIALQNNCKMGGELMRVDIPLKLVMI 629
Qy      544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOALRAMNSC 603
Db      630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOALRAMNSC 689
Qy      604 NEYMSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRF 663
Db      690 NEYMSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRF 749
Qy      664 AOSGRLQNPPLPGTVIDVETREPWYDFIVSQAVRSQSVSPTHYNYVYDNGSLKPDHIQ 723
Db      750 AOSGRLQNPPLPGTVIDVETREPWYDFIVSQAVRSQSVSPTHYNYVYDNGSLKPDHIQ 809
Qy      724 RLTYKLCIHYNNMPGVIKVPAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 775
Db      810 RLTYKLCIHYNNMPGVIKVPAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 861

RESULT 3
AA190234
ID      AA190234 standard; protein; 862 AA.
XX      AA190234;
XX      29-AUG-2000 (first entry)
XX      Mouse piwi protein, designated miwi.
XX      DE
XX      piwi family protein; miwi; miwi; gene therapy; tissue dystrophy;
XX      anaemia; immunodeficiency; male infertility; mouse; ds.
XX      OS
XX      Mus sp.
XX      Location/Qualifiers
FH      Key
FT      Misc-difference 90
FT      /label= Leu, Ile
FT      /note= "encoded by NTT"
FT      Misc-difference 216
FT      /note= "unspecified amino acid; encoded by NTC"
FT      Misc-difference 383
FT      /label= Leu, Ile
FT      /note= "encoded by NTC"
FT      Misc-difference 816
FT      /label= Leu, Ile
FT      /note= "encoded by NTC"
XX      MO200032039-A1.
XX      PD
XX      08-JUN-2000.
XX      03-DEC-1999; 99WO-US028764.
XX      PR      04-DEC-1998; 98US-0110901P.
XX      PA      (UYDU-) UNIV DUKE.
XX      Lin H;
XX      PI
XX      DR      WPI; 2000-412085/35.

```

```

DR      N-PSDB; AAA07587.
XX      PT      piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX      therapy of diseases such as cancer and in various research and diagnostic
XX      applications.
XX      PS      Claim 4; Page 180-185; 201pp; English.
XX      CC      This sequence represents the mouse piwi family protein, designated miwi.
XX      CC      The piwi family nucleic acids and polypeptides are used in gene therapy
XX      CC      of diseases such as cancer and also in various research and diagnostic
XX      CC      applications. The sequences can also be used to treat tissue dystrophy,
XX      CC      anaemia, immunodeficiency, and male infertility
XX      SQ      Sequence 862 AA;
Qy      Query Match      95.5%; Score 3957; DB 3; Length 862;
Qy      Best Local Similarity 95.9%; Pred. No. 0; Indels 0; Gaps 0;
Qy      Matches 740; Conservative 19; Mismatches 13;
Qy      4 GYNTRONLDHYVESKTSGGIIIVRLSTNHFRLTSRPPQALYQYHIDVNPIMBARLRBAL 63
Qy      91 GYNTRONLDHYVESKTSGGIIIVKLSNHFRLTSRPPQALYQYHIDVNPIMBARLRBAL 150
Qy      64 LFQHEDLIGKCHAFDGTLLFLPKLQKQVTEVFSKTRGQEDRITLTLNLPPTSPTCL 123
Qy      151 LFQHEDLIGKCHAFDGTLLFLPKLQKQVTEVFSKTRGQEDRITLTLNLPPTSPTCL 210
Qy      124 QFYNIIFRLIKIMLQOIGRYNYNPNDPIDPSRLVWPQFTTSLIOYENSIMLCTDV 183
Qy      211 QFYNIIFRLIKIMLQOIGRYNYNPNDPIDPSRLVWPQFTTSLIOYENSIMLCTDV 270
Qy      184 SHKVLRSFTVDLPNENFHHQTEHKKFQEOVSKEELIGLVLTIKNNKYTRVDIDMDQNK 243
Qy      271 SHKVLRSFTVDLPNENFHHQTEHKKFQEOVSKEELIGLVLTIKNNKYTRVDIDMDQNK 330
Qy      244 STEKKADGSEVSEFLFYRKQYQCEITDLPQVLPVQPKRRRPGGTLPGPAMILPELCYL 303
Qy      331 STEKKADGSEVSEFLFYRKQYQCEITDLPQVLPVQPKRRRPGGTLPGPAMILPELCYL 390
Qy      304 TGLTDMKNDNFVMDLAHVRFLTPBQRORFVGLIDYIHKNDVQRELRDMLSDSNL 363
Qy      391 TGLTDMKNDNFVMDLAHVRFLTPBQRORFVGLIDYIHKNDVQRELRDMLSDSNL 450
Qy      364 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 423
Qy      451 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 510
Qy      424 SLIQNLFKVTPAMGMQMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 483
Qy      511 SLIQNLFKVTPAMGMQMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 570
Qy      484 AIKKYLTCTDPTSCQCVARTLGKQOTVMAITKIALQNNCKMGGELMRVDIPLKLVMI 543
Qy      571 AIKKYLTCTDPTSCQCVARTLGKQOTVMAITKIALQNNCKMGGELMRVDIPLKLVMI 630
Qy      544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOALRAMNSC 603
Qy      631 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOALRAMNSC 690
Qy      604 NEYMSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRF 663
Qy      691 NEYMSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRF 750
Qy      664 AOSGRLQNPPLPGTVIDVETREPWYDFIVSQAVRSQSVSPTHYNYVYDNGSLKPDHIQ 723
Qy      751 AOSGRLQNPPLPGTVIDVETREPWYDFIVSQAVRSQSVSPTHYNYVYDNGSLKPDHIQ 810
Qy      724 RLTYKLCIHYNNMPGVIKVPAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 775
Qy      811 RLTYKLCIHYNNMPGVIKVPAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 862

```

RESULT 4
AD65285
ID ADB65285 standard; protein; 829 AA.
XX
AC ADB65285;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone TEST120060830.
XX
KW Human, pharmaceutical, diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX
OS Homo sapiens.
XX
PN EPI308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
XX
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuno Y;
XX
DR WPI; 2003-450961/43.
XX
DR N-PSDB; ADB63315.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease transduction-related proteins, and
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 829 AA;

Query Match 94.7%; Score 3883; DB 7; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 733; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	GVNTRQNDHYVESKTSAGSIIVRLSTNHPFLTSRPMQALVQYHIDVNPFLMEARRLSAL	63
DB	90	GVNTRQNDHYVESKTSAGSIIVRLSTNHPFLTSRPMQALVQYHIDVNPFLMEARRLSAL	149
QY	64	LPQHEDLIGKCHAPDGTLLFLPKRLQKQVTEVFSKTRNGEDVRITITVLTNLPPTSPTCL	123
DB	150	LPQHEDLIGKCHAPDGTLLFLPKRLQKQVTEVFSKTRNGEDVRITITVLTNLPPTSPTCL	209
QY	124	QFYNIIFPRLKINMLQIGRNNYNNPDIDIPSHRLVYMPGFTSLIQYNSIMLCTDV	183
DB	210	QFYNIIFPRLKINMLQIGRNNYNNPDIDIPSHRLVYMPGFTSLIQYNSIMLCTDV	269
QY	184	SHKVLRSSTVLDFMENFPHQTEHKEHFOVSKEELIGLVLTIRYNNKYRVYVDIDMDQNPX	243
DB	270	SHKVLRSSTVLDFMENFPHQTEHKEHFOVSKEELIGLVLTIRYNNKYRVYVDIDMDQNPX	329
QY	244	STFKKADSEVSFLFYKQYNOETDLKQVYVQSPRRRPGGTLPGPAMLIPELCYL	303
DB	330	STFKKADSEVSFLFYKQYNOETDLKQVYVQSPRRRPGGTLPGPAMLIPELCYL	389
QY	304	TGLTDKXNDPNNVMDLAVHTRLTPEQFORSEVGRLLIDYHKNQVQRELRDWSGSPDSNL	363
DB	390	TGLTDKXNDPNNVMDLAVHTRLTPEQFORSEVGRLLIDYHKNQVQRELRDWSGSPDSNL	449
QY	364	LSFSGRLLIQTEKIHQGGKTFDYNPOPADWSKETGAPLISVKPDLNMLITRREYKAN	423
DB	450	LSFSGRLLIQTEKIHQGGKTFDYNPOPADWSKETGAPLISVKPDLNMLITRREYKAN	509
QY	424	SLIQNLFVYTPAMGQMKAMIEVDRETEAVLVLOQKVTADTQIVVCLLSNRKQYD	483
DB	510	SLIQNLFVYTPAMGQMKAMIEVDRETEAVLVLOQKVTADTQIVVCLLSNRKQYD	569
QY	484	AIKTYLCTDCTPSCQVARTLQKQYMAIATIALQMKRGGELRVDPILKLVITV	543
DB	570	AIKTYLCTDCTPSCQVARTLQKQYMAIATIALQMKRGGELRVDPILKLVITV	629
QY	544	GIDCHDMTAAARRSIAGVASINSGMTRFRCIFODRGSELVGLKVCLOALRAMNSC	603
DB	630	GIDCHDMTAAARRSIAGVASINSGMTRFRCIFODRGSELVGLKVCLOALRAMNSC	689
QY	604	NEWPSRIIVRDGVGDQLKTLVNYEVPFLDCLKSIIGRYNRLTYIVVKKVNTRE	663
DB	690	NEWPSRIIVRDGVGDQLKTLVNYEVPFLDCLKSIIGRYNRLTYIVVKKVNTRE	749
QY	664	AQSGRLQNPUGTYIVDEVTRPPEWYDFIVSAVRSGSVPTRYNYTYDNGSKPDHIQ	723
DB	750	AQSGRLQNPUGTYIVDEVTRPPEWYDFIVSAVRSGSVPTRYNYTYDNGSKPDHIQ	809
QY	724	RLTYVLCIIYNNWP 737	
DB	810	RLTYVLCIIYNNWP 823	

RESULT 5
ABP67995
ID ABP67995 standard; protein; 523 AA.
XX
AC ABP67995;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related polypeptide SEQ ID NO 2603.
XX
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US043704.

PR 20-NOV-2000; 2000US-0252222P.
 PR 06-FEB-2001; 2001US-0267011P.
 PR 28-MAR-2001; 2001US-0279670P.
 PR 10-JUL-2001; 2001US-0304037P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretist H, King GB;
 XX MPI: 2002-608400/65.
 DR N-PSDB; ABV89287.
 XX
 PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer.
 XX
 PS Claim 2, SEQ ID NO 2603; 266pp + Sequence listing; English.
 XX
 CC The invention relates to a human colon tumor expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 523 AA;

Query Match 67.4%; Score 2764; DB 5; Length 523;
 Best Local Similarity 100.0%; Pred. No. 1.7e-260;
 Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 VSFLEYRKNQNOETIDLKQPVVQPKRRRGGTLPQAPMLPELCYLTGLTDKMD 313
 DB 2 VSFLEYRKNQNOETIDLKQPVVQPKRRRGGTLPQAPMLPELCYLTGLTDKMD 61
 QY 314 FVWMDLAVHTLTTEOROREVGRLLIDYHKNDYORLEBDMGSPDSNLSFSGRIQT 373
 DB 62 FVWMDLAVHTLTTEOROREVGRLLIDYHKNDYORLEBDMGSPDSNLSFSGRIQT 121
 QY 374 EKHGGKTFDYNPQFADWSKETRGAPLISVPLDNMLIYTRNVEANSLIQNTFRYT 433
 DB 122 EKHGGKTFDYNPQFADWSKETRGAPLISVPLDNMLIYTRNVEANSLIQNTFRYT 181
 QY 434 PAMGOMKRAIMIEVDRTREATLRLQOKTADTQIVVCLSSNKKDYDAIKKYLCDG 493
 DB 182 PAMGOMKRAIMIEVDRTREATLRLQOKTADTQIVVCLSSNKKDYDAIKKYLCDG 241
 QY 494 PPSOCVAVARTLGKQQTWAIATKIALQNCMCGELWEVDIPLKLVIVGIDCYHDMTA 553
 DB 242 PPSOCVAVARTLGKQQTWAIATKIALQNCMCGELWEVDIPLKLVIVGIDCYHDMTA 301
 QY 554 GRRSIAGPVASINEGMTWFSRCIFQDRGQELVDGLKVCLOALFANNSCNEIWPRIIV 613
 DB 302 GRRSIAGPVASINEGMTWFSRCIFQDRGQELVDGLKVCLOALFANNSCNEIWPRIIV 361
 QY 614 YVDGVDGGLKTLVAVYVPOFADCKSIGRGNPRLTYIVVKKRNTVFPAFSGGRLONP 673
 DB 362 YVDGVDGGLKTLVAVYVPOFADCKSIGRGNPRLTYIVVKKRNTVFPAFSGGRLONP 421
 QY 674 LFGTVIVDEVTREPVEYDFTVSQAIVSSGVSPTNVIYDNGSLKPHDIQRLTYLCHY 733
 DB 422 LFGTVIVDEVTREPVEYDFTVSQAIVSSGVSPTNVIYDNGSLKPHDIQRLTYLCHY 481
 QY 734 YNMPGVIRVAPCOYAHKLAFLVGSIHREPLSTSNLTYL 775
 DB 482 YNMPGVIRVAPCOYAHKLAFLVGSIHREPLSTSNLTYL 523

RESULT 6
 ABU61933
 ID ABU61933 standard; protein; 523 AA.
 XX
 AC ABU61933;
 XX
 DT 19-AUG-2003 (first entry)
 XX
 DE Human Hiwi protein.
 XX
 KW Human; Hiwi-Hy; stem cell maintenance factor; immunosuppressive;
 KW antidiabetic; antiparkinsonian; cerebroprotective; neuroprotective;
 KW diabetes; CNS disease; graft-versus-host disease; Parkinson's disease;
 KW organ transplantation; biological-based sensor; Hiwi.
 XX
 OS Homo sapiens.
 XX
 PN US2003027255-A1.
 XX
 PD 06-FEB-2003.
 XX
 PF 26-OCT-2001; 2001US-00029495.
 XX
 PR 20-AUG-1999; 99US-00378667.
 XX
 PA (BALI/) BALINGER D G.
 PA (BRMA/) DRMANAC R T.
 PA (LABA/) LABAT I.
 PA (STRC/) STACHE-CRAIN B.
 PA (DICK/) DICKSON M C.
 PA (JONE/) JONES L W.
 PA (XUEA/) XUE A.
 PA (TANG/) TANG Y T.
 PA (LIUC/) LIU C.
 PA (ASUN/) ASUNDI V.
 XX
 PI Ballinger DG, Drmanac RT, Labat I, Stache-Crain B, Dickson MC;
 PI Jones LW, Xue A, Tang YT, Liu C, Asundi V;
 XX
 DR MPI: 2003-479536/45.
 XX
 PT New isolated stem cell maintenance factor polynucleotide and polypeptide,
 PT useful for the diagnosis and treatment of disorders such as diabetes, CNS
 PT diseases, graft versus host disease, Parkinson's disease and organ
 PT transplantation.
 XX
 PS Disclosure; Fig 1; 63pp; English.
 XX
 CC The invention relates to a new isolated polynucleotide comprising: (a) the
 CC human partial Hiwi-Hy cDNA appearing as A062541; (b) having the stem
 CC cell maintenance factor protein coding nucleotide sequence of a
 CC polynucleotide of (a); or (c) having the mature stem cell maintenance
 CC factor protein coding nucleotide sequence of (a). Also include are an
 CC isolated polynucleotide encoding a polypeptide with stem cell maintenance
 CC factor activity that hybridizes under stringent conditions to the
 CC complement of (or has a sequence identity greater than 90% to A062541);
 CC an expression vector comprising the Hiwi-Hy cDNA, a host cell genetically
 CC engineered to contain or express the Hiwi-Hy cDNA, in operative
 CC association with a regulatory sequence that controls expression of the
 CC polynucleotide in the host cell, the partial Hiwi-Hy stem cell
 CC maintenance factor appearing as ABU61932, an anti-Hiwi-Hy antibody,
 CC detecting Hiwi-Hy in a sample (comprising contacting the sample with a
 CC compound that binds to and form a complex with the polypeptide),
 CC identifying the binding compound and modulating cell survival and/or
 CC death (comprising contacting the cell with Hiwi-Hy for a time sufficient
 CC to modulate the activity). The methods and compositions of the present
 CC invention are useful for the diagnosis and treatment of disorders such as
 CC diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease
 CC and organ transplantation. They can also be used for re-engineering
 CC damaged or diseased tissues, manufacture of pharmaceuticals and in the
 CC development of biological-based sensors. The present sequence is a

CC homologous sequence to huma Hiwi-Hy
XX
SQ Sequence 523 AA;

Query Match 67.4%; Score 2764; DB 7; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.7e-260; Indels 0; Gaps 0;
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 254 VSFLEVRKQYNQDEITDLKQPVLSQPKRRRPGGTLPGPAMLIPELICYLTGLTDKMD 313
DB 2 VSFLEVRKQYNQDEITDLKQPVLSQPKRRRPGGTLPGPAMLIPELICYLTGLTDKMD 61
CC 314 FNNWKDLAVHTRLPEQRQREVERGLIDYIHKNDVQRELDWGLSFSNLSFGRL 373
DB 62 FNNWKDLAVHTRLPEQRQREVERGLIDYIHKNDVQRELDWGLSFSNLSFGRL 121
CC 374 EKHGGKTPDYNPOFADMSKETRGAPLISVPLDNLITYSRYEAANSLIQNLFKY 433
DB 122 EKHGGKTPDYNPOFADMSKETRGAPLISVPLDNLITYSRYEAANSLIQNLFKY 181
CC 434 PAMGOMKALMIEVDRTETAYLAVLQOKTADTQIVVCLLSNRKDKYALKYCTDC 493
DB 182 PAMGOMKALMIEVDRTETAYLAVLQOKTADTQIVVCLLSNRKDKYALKYCTDC 241
CC 494 PTPGQCVARTLGRQOTMAATKIALQNMCKMGGELMRVDIPKLVMIVGIDCYHDM 553
DB 242 PTPGQCVARTLGRQOTMAATKIALQNMCKMGGELMRVDIPKLVMIVGIDCYHDM 301
CC 554 GRSIAGFVASINQMTWFSRCIFODRGELVDGLKVCLOALFANNSCNEYMPRIY 613
DB 302 GRSIAGFVASINQMTWFSRCIFODRGELVDGLKVCLOALFANNSCNEYMPRIY 361
CC 614 YRDVGSGQLKTLVNEVPOFLDKSIGGVNRLTVIYKKRVNRPFAOSGRLQNP 673
DB 362 YRDVGSGQLKTLVNEVPOFLDKSIGGVNRLTVIYKKRVNRPFAOSGRLQNP 421
CC 674 LFGTVIDVETRPMEYDFIIVSAVRSGSVPTHYNTIYNSGKPDHIOQLTYKCH 733
DB 422 LFGTVIDVETRPMEYDFIIVSAVRSGSVPTHYNTIYNSGKPDHIOQLTYKCH 481
CC 734 YNMGVIRVPAPCOYAKTLAVGOSTHREPNLSLRVYL 775
DB 482 YNMGVIRVPAPCOYAKTLAVGOSTHREPNLSLRVYL 523

RESULT 7
ADB65492
ID ADB65492 standard; protein; 852 AA.
XX
AC ADB65492;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone TEST120255460.
XX
KM Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KM cell regeneration; membrane protein; signal transduction-related protein;
KM transcription-related protein; osteoporosis; neurological disease;
KM cancer; tumour.
XX
XX Homo sapiens.
XX OS
XX EN EPI308459-A2.
XX
XX PD 07-MAY-2003.
XX
XX PF 28-MAR-2002; 2002EP-00007401.
XX
XX PR 05-NOV-2001; 2001JP-00379298.
XX
XX PR 25-JAN-2002; 2002US-00350978.
XX
XX PA (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto Y, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuo Y;
XX
DR MPI: 2003-450961/43.
XX
DR N-PSDB; ADB63522.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1, Page: 222pp; English.

CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

XX
SQ Sequence 852 AA;

Query Match 51.2%; Score 2101.5; DB 7; Length 852;
Best Local Similarity 51.0%; Pred. No. 2.2e-195;
Matches 397; Conservative 148; Mismatches 224; Indels 9; Gaps 4;

CC 1 MIFGVNTRQNLDPKESKTSGGIIVRLSTNHFRLTSRPPQALYQYHIDNPLMARLR 60
DB 81 MDSICTREKLAHVNNCKTSSGIPVKLVNLFNDFQDQOLYQYHVTYIPDLASRRR 140
CC 61 SALTFQHEDLICKAFAGTTLFELPKRLQOKTVEVSKTRNGEDVRIITLNEAPPTSP 120
DB 141 IALLYSHSELSENAKAFDAILFSLQKLEKYTELSSETGRTETKMTTLKRELPSSSP 200
CC 121 TCIQFNIIFRRLKIMNLOQIGRNYNPNPIDIPSHRLVWPGFTTSIIQYENSIMC 180
DB 201 VCIQVFNIFRILKLSWYQIGRNFYNPSEMEIPQKLSIMPFAISVSYFERKLIFS 260
CC 181 TVSHKVLASEVVDPMFRFHTQTEBHFQOEVSLELGLVLTLYNNKTVRVDDIDMQ 240
DB 261 ADVSYKVLNENFVLEFMTALCQRTGLSCTQYCEKQLGLVLTLYNNKTVSIDIIDSV 320
CC 241 NPKSTFKKADSGSEVFELEVRKQYNQDEITDLKQPVLSQPKRRRPGGTLPGPAMLIPEL 300
DB 321 KPTHFQKRDGRLIYVDYKQYDITVSDNLQPTVSLKKKRDNSEAQ-LALILPEL 379
CC 301 CYLTGLTDMKRDVNMKDLAVHTRLPEQRQREVERGLIDYIHKNDVQRELDWGLSG 360
DB 380 CELTGLTDQATSDFOLMRAVAKTRLSBGRQRLARLVNDIQRTNARFELLETGLHFG 439
CC 361 SNLSFSGRIIQTEKIHQGGKTFDYNPQ---FADMSKETRGAPLISVPLDNLITYR 417
DB 440 SQ-ISLTGIVPSEKIMQ---DHICQVSAADMSKDIRCTKINAGSLNMTWLILCSDR 494

QY 418 NTEANSIQLNLFKVTTPAMQMRKAIMIEVDRTAELRVLOQKVTADTQIVCLISN 477
 DB 495 TEYVASEFNLCLRRVAGSGWVNDYPIKIKVQENPAFAFRAIQYVVDVQLVWCLIPSN 554
 QY 478 RDKXDAIKKYLCTCOPPSQCVARTLGKQOTVNAIKTALQNMCKMGEIWRVIDPL 537
 DB 555 QRTYDSIKKILSSDCEVPSQCLARTKQNMGSIAIKTKAMQMTCKLGGELMAVELPL 614
 QY 538 KLVMTVGIDCYHDMTAGRSISAGFVASINEGWTMFRSCIFQDRQGLVLDGKYCLQAL 597
 DB 615 KSLMTVGIDCYKDALSKDMVWVGVASVNPRTMFRSCIIQRTMTVDADCLKFMGAL 674
 QY 598 RAMNSCNEMBSRIIVYRDVGDGQKTLVNTVEVQPLDCLKSGIRGYNPLTVIVKCR 657
 DB 675 NKMYKNHDLPRRIIVYRAGVGDGQLETLIEYEVQQLSSVAESQSNSSRLSVIVRKK 734
 QY 658 VNTREFAQSGGRLQNPFGTVIDVEVTRPEWYDFPIVSOAVRSGSPPTHNVVYDNGSL 717
 DB 735 CMRPFTEMRNTVQNPGLCTIVVDSATRNEMIDFYLISQVACRGTVSTYNNVYIDNGL 794
 QY 718 KPDHIOQLTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
 DB 795 KPDHIOQLTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 852

RESULT 8
 ABB62102
 ID ABB62102 standard; protein; 866 AA.
 AC ABB62102;
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 13098.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN MO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 FR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PERK) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW,
 DR WPI; 2001-6556860/75.
 DR N-PSDB; ABL06205.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU10511), expressed DNA
 CC sequences (ABU10840-ABU16175) and the encoded proteins (ABU57737-
 CC ABU72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences

SO Sequence 866 AA;
 Query Match 35.8%; Score 1468; DB 4; Length 866;
 Best Local Similarity 39.4%; Pred. No. 2e-133;
 Matches 305; Conservative 154; Mismatches 291; Indels 24; Gaps 12;

QY 17 SKTSGSGLIIVRLSTNFRILTRSPOMALYQYHIDYNLMEARLSALFQHEDLIGKCHA 76
 DB 102 SKGVGVGTHITVQANFKYAKRPNMTIYQRYVDFTDVAETRLRRFLVFKHLLGG-YI 160
 QY 77 FQGTILFLP--KRLQOK--VTEVSKTRNGEDVIRITITLNEPLPTSPCLQFVNIIFR 131
 DB 161 FQGTINFCINQFAVODSPVYELVTKSRAGENIEIKIYAVGSQSTDAEQFQVNLILIR 220
 QY 132 RLKINMLQOIGANNYNPNPDIDIPSHRVIMVGFSTISLOVNSIMLCQDVSHKYLRS 191
 DB 221 RAMEGDLKLVSYTYDDPOAKINLENFRQQLNFGVYTSIRQENIDILLSCICHKMYRT 280
 QY 192 TVLDFMNFYHQTEHKFQEVSKELLGLVLTLYNNKTYRVDIDMDQNPXSTFKKADG 251
 DB 281 TLVNIISDAIRSDS--YQSTFKRAVMGVILTDYNNKTYRIDVDVFGSTPLCKFKTNDG 338
 QY 252 SEVSFELEYRKYQNOEITDQKQVLMASQPKRRGPGTILPGAMLIPELCYLGLTKDKR 311
 DB 339 -EISYDYTKKRYNIIIRDLKQPLVMSRPTDKVIRGN--DQALMITIPELARATGMIDNR 396
 QY 312 NDFNVKDLAVHTRLTPEQRQREVGRLIDYIHKNQVQRELRDWGISFDSNLSFSGRL 371
 DB 397 ADFRTLRAVSEHTRAMPDRIRIELRMFNRLKSCQSVETLKSWMIEELSALVEIPAVTL 456
 QY 372 QTEKIHQGGKTDPYNQPADNSKETRGALIVSKPLDMWLLYTRANYAANSIQLNLR 431
 DB 457 PPEKILFNGQKIFVCCDARDMTNBERPTCSMFKNVIRNRYVITPSSNLTRETOEFVQMCIR 516
 QY 432 VTPAMQMRKAIMIEV--DRTAELRVLOQKVTADTQIVCLISNRKQKXDAIKKYC 490
 DB 517 TASSMGNICNPIYEBIPDDRNGTYSQALDNAAANDPQIVMVMSSPNEKXSCIKKRC 576
 QY 491 TDCPTPSQCVARTLGKQOT---VNAIATKIALQNMCKMGEIWRVIDPLKLVMTVGID 546
 DB 577 VDRPVPQVYTLKVIAFRQCKPTGLMSIATKVIOAKMKAAPQVVIPLHGLMTVGSD 636
 QY 547 CYHDMTAGRSISAGFVASINEGWT--RMPFRSCIFQDRQGLVLDGKYCLQALRAMNSCN 605
 DB 637 VCHSPKNKXSYGAVYATMDQKESFRYTSVNEHITKQGLSEQMSVNAACALRSYQBOCR 696
 QY 606 YMPSRIVYRDVGDGQKTLVNTVEVQPLDCL---KSIRGYNPLTVIVKCRVNTR 661
 DB 697 SLPERILFPRDVGQGLYQVNVSEVNTLKDLDEIYKSAGKQEGCRAFTFIIVSKRINR 756
 QY 662 FFAQSGGRLQNPFGTVIDVEVTRPEWYDFPIVSOAVRSGSPPTHNVVYDNGSLKDPH 721
 DB 757 YF--TGRH--NVPGVVDDVITLPERYDFVLSQAVRIGVSPSYNISDMGMLADK 812
 QY 722 IQRITTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
 DB 813 LQMLSYKQTHMYNNSGTIRVAVCHYAHKLAFLVABSINRAPSGLQNGQYFL 866

RESULT 9
 ABU61934
 ID ABU61934 standard; protein; 866 AA.
 AC ABU61934;
 XX
 DT 19-AUG-2003 (first entry)
 XX
 DE Fruitfly Aubergrine protein.
 XX
 KW Hw1-HY; stem cell maintenance factor; immunosuppressive; antidiabetic;
 KW antiparkinsonian; cerebroprotective; neuroprotective; diabetes;
 KW CNS disease; graft-versus-host disease; Parkinson's disease;
 KW organ transplantation; biological-based sensor; fruitfly; aubergrine.

PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
PA (XUEA/) XUE A.
PA (TANG/) TANG Y T.
PA (LIUC/) LIU C.
PA (ASUN/) ASUNDI V.
XX
PI Ballinger DG, Drmanac RT, Labat I, Stache-Crain B, Dickson MC;
PI Jones LW, Xue A, Tang YT, Liu C, Asundi V;
XX
DR WPI; 2003-479536/45.
XX
XX
XX New isolated stem cell maintenance factor polynucleotide and polypeptide,
PT useful for the diagnosis and treatment of disorders such as diabetes, CNS
PT diseases, graft versus host disease, Parkinson's disease and organ
PT transplantation.

PS Disclosure; Fig 1; 63pp; English.

XX
CC The invention relates to a new isolated polynucleotide comprises: (a) the
CC human partial Hiw1-Hy cDNA appearing as ACG62541; (b) having the stem
CC cell maintenance factor protein coding nucleotide sequence of a
CC polynucleotide of (a); or (c) having the mature stem cell maintenance
CC factor protein coding nucleotide sequence of (a). Also include are an
CC isolated polynucleotide encoding a polypeptide with stem cell maintenance
CC factor activity that hybridises under stringent conditions to the
CC complement of (or has a sequence identity greater than 90% to ACG62541),
CC an expression vector comprising the Hiw1-Hy cDNA, a host cell genetically
CC engineered to contain or express the Hiw1-Hy cDNA in operative
CC association with a regulatory sequence that controls expression of the
CC polynucleotide in the host cell, the partial Hiw1-Hy stem cell
CC maintenance factor appearing as ACG61932, an anti-Hiw1-Hy antibody,
CC detecting Hiw1-Hy in a sample (comprising contacting the sample with a
CC compound that binds to and form a complex with the polypeptide),
CC identifying the binding compound and modulating cell survival and/or
CC death (comprising contacting the cell with Hiw1-Hy for a time sufficient
CC to modulate the activity). The methods and compositions of the present
CC invention are useful for the diagnosis and treatment of disorders such as
CC diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease
CC and organ transplantation. They can also be used for re-engineering
CC damaged or diseased tissues, manufacture of pharmaceuticals and in the
CC development of biological-based sensors. The present sequence is a
CC homologous sequence to human Hiw1-Hy

XX
SQ Sequence 844 Aa;

SQ Sequence 844 AA;

Query Match: 35.0%; Score 1434; DB 7; Length 844;
 Best Local Similarity: 38.0%; Pred. No. 4e-130;
 Matches 294; Conservative 16; Mismatches 286; Indels 32; Gaps 13

```

QY 17 SKTSGSGIVRSTHSTHFFLTSPQALVOYHIDYNPMBEARLREALLFOHEDDILGCKHA 76
Db 88 SKKGIDGVPMLOIWFPRFKTCEPMRI VHYHVEFSPSIENPRVNGVLSNEANILGSGYL 147
QY 77 FDGTLLEPKLOQCVTEVESKTRNGEDVYRITILTNELPSPSPCLOFYNIIPRLKI 136
Db 148 FDGLQFTTRKEBQITLVGSKSKDIEKISIKVGVFSCEBPFLOYANILIRSMKG 207
QY 137 MNLQOIGNNYNPNPIDIPSHRLVIMPGFTTSILOYSIMLCTDVBHKLVRSETVDF 196
Db 208 INLEIVGNLLEDPRAKIEIRREFKMLMPYEYSISIOHEKDILGIEITHKVRRTIYDI 267
QY 197 MENFQOTEBEHFQEOVSKEILIGLVULTYXNNKTRVVDIDWQDNPKSTFFKADSGEVSF 256
Db 268 MRCSCHNPABH--QDEVRNNVLDLIVLEFNNRTRRINDVDVGQTPKSTF--SCKARDISF 324
QY 257 LEYVAKQNOEITDELKOPVLVSOQPRRRPGGTLGCPAMLIPELCYLTGLTDMENDENVY 316
Db 325 VEYITLKINIRIRRDHNOPLLIK--NRDKALKNMASEVLVYLIELCRVYGLNANEMSNPQL 383
QY 317 MKDLAVHRLRLTEQOREVGLRIDYIHKDNDQ--RELMDGSLSDSNLSFSGRILOT 373

```

```

Db 384 MRAMSSYTBMMKOR---TDBLRANHRLQNTPEBZYVLDDMMHEDLNKTEVOGRILGO 440
Qy 374 EKI--HQGKRTDYNPQPADMSKEXRGNPLSYVP--LDMMLLYTBRYEANSILQNL 429
Db 441 QNTVFVNGVAVAGEN--ADMQRHFRDORMLTTPSDGIDBRNAVAPRASHELRTILDSL 437
Qy 430 FKYPYPMGMQMKR-AIMIEVDRTAEAYURLVQOKVTADTOICVLLSSNRKDYAIKUY 468
Db 498 YRASGSMGRISPOEFIIYDRJGTYYRAMDDCVASDPILCLTGVENDNAERYSIKRA 557
Qy 489 LCTDCPTBPQCVANTLKGQOVTALATKIALQANCKMKGSELRVDPILYLTWITGDCY 548
Db 558 GYDRAVPPQVAVTELTITKPKPYSLSMIARKIAIQNCCKLGTTPWMIELPISGLMTIGDIA 617
Qy 549 HDWTAQRSLAGVASIN-BGTMFRFSRCIPDRGQELVJGCKLOALRANMSCEYUW 607
Db 618 KSTRDKRKRIYAGALLISMDCQNSITYFSTVTECSAFDYLTANTLWMIMAKALROYOHENRKU 677
Qy 608 PSRIIVYRDGVGDQOLKTLVNYEYRQFLDCKSIGRGY-----NPLTYIYVKKRQVNR 661
Db 678 PSRIYVYRDGVSGSLKQLFEBFEVDITIEKTL--EYARQULSPPOAIYVTRSMNTR 724
Qy 662 FFMQSGRLQNTLPBGVIVDEYTRBEMDFEIVQANRSGSVSTHNVYLYDMSGKPDH 721
Db 735 FFLNNG---QNPPEGLIVDVTTLPERDFEVLVQOVRQGVSTSYNVLYSSMGLSPEK 750
Qy 722 IORTLYTKLGHIIYNNPQGVIRVAPACQVYAHKLAFLVGOGSIHREPMILSNRLYYL 775
Db 791 MQULTYIKMGHLYNMSGTRVPAYVCQYAKLATALVGNTHSHIPGNALKEKFFYYL 844

```

RESULT 11
ADP94999

ID ADB94998 standard; protein; 844 AA

AC ADB94998

DT 04-DEC-2003 (first entry)

DE Fruitfly stem cell protein, piwi.

Frutitfly; haematopoietic growth regulatory protein; HIWI; cytosstatic; KW

CD34+ haematopoietic stem cell; leukaemia; cancer; piwi.

Drosophila melanogaster

PN US2003129599-A1.

PD 10-JUL-2003

PF 10-JAN-2002; 2002US-00043774.

PR 10-JAN-2002; 2002US-00043774.

PA (SHAR/) SHARMA A.

XX
XX
Chomsky, J. Professor

XX
WB1: 2003-74434E/70

Now, let's just go ahead and

PT useful as cancer therapeutic.

Example 2; Fig 1; 39pp; English

CC The invention relates to a nucleic acid comprising a nucleotide sequence
CC encoding a human hwi1 protein (a haematopoietic growth regulatory gene).
CC Also included are the Hwi1 protein, the Hwi1 cDNA (used as a probe), a
CC recombinant expression construct comprising the Hwi1 cDNA, a cell culture
CC transformed with the recombinant expression construct, screening a
CC compound for modulating human hwi1 gene activity in cells expressing the

human hiwi gene product, a cell membrane preparation (or cytosolic preparation or nuclear preparation) comprising a human hiwi gene product (or derivative) and identifying a compound that induces or increases hiwi gene expression in mammalian cells. Culturing bone marrow or peripheral blood culture in the presence of compounds identified by the methods is useful for increasing retention of primitive CD34⁺ haematopoietic stem cells. Alternatively the method comprises introducing a retroviral or lentiviral expression construct into the cells. The compounds are useful for haematopoietic stem cell development and maintenance and for inhibiting leukaemia cell growth and treating cancer. The present sequence represents the fruitfly piwi protein, a homologue of the Hiwi protein.

Sequence 844 AA;

Query Match 35.0%; Score 1434; DB 7; Length 844;

Best Local Similarity 38.0%; Pred. No. 4e-130; Mismatches 294; Conservative 162; Mismatches 286; Indels 32; Gaps 13;

17 SKTSSGIIYRLSTNHRFLTSRPMALYOHIDYMLMEARLRSLFQHEDLIGKCHA 76
 88 SKKGTGVPVWLQTNFRLKTPKEMRIYHYHFEPSIENPRVGVLSNHNLLGSGYL 147
 77 FDGILFLPKLQOKYTEVFSKTRNGEDVRIITITLNLPTSPCLQFYNIIFRLIKI 136
 148 FDGLQFLTRKFEQETIVLSGSKLDIEYKISIKFVGFISCAEPFLQVNLILRRSMKG 207
 137 KNLQOIGRNYVNPNDIPSHRLVYVPGFTSIIQYENSIMLCYDVSFKVLRSETVDF 196
 208 LNLVLGRNLPDPRAKIEIRFPMELMPYETSIRQHEKDIILGTETIKVMTETIYI 267
 197 MENFYQTEBHKQEOVSKEILGLVYLTKNKTYRVDIDWDQNPSTFKKADGSEVSF 256
 268 MRCSNPARH--QDEVKRVNVDLIVLTIDNNRTYRINDVDFQGTFRSTF--SCKGRDISF 324
 257 LEYRKQYNOEITDLKQPVIVSOPKRRRGGTLPEPAMLIPELCYITGIDTMRDNFV 316
 325 VEYLTQYINRIRHNPPLISK--NRDKALKTNASELVAVIPELCRTGLTAMRNENFOL 383
 317 MKDLAVTRLTPEBROREVRGLDIYHKNNVQ---RELADGLSPSNLSPSGRILO 373
 384 MRAVSTYRMPKOR---TDRLRAFNRLOQTESVAVLDMWELDXVTEVQSKIIQ 440
 374 EKI--HOGGKTFDYNPOFADWSKETRGAPLISVP--LDWMLIYTRRYEANSILQNL 429
 441 QNIVFHNGKVPAGN---ADWQRHFRDQMLTTPSDGLDRAVAVIAPQRNHEHLERTLDSL 497
 430 FKVTTPANGMQRK-AIMEVDDRTREALVRLQOKVYATDQIVVCLLSNNKDKYDAIKY 488
 498 YRASGNGMLRIRSDQEFITIDRTGYVRAMDCVRSDPKIILCLVNDNAERYSSIKR 557
 489 LCTGCPFPSCVAVATLIGKQOTVMAITKIALQNCNKGELMEVDIPLKLVMTIGDY 548
 558 GYVRAVFTQVYVLTKTTPKPYSLMSIAIKIAIQNCKLGYTPMIEIPLSGIMTIGDIA 617
 549 HDMTAGRSIAGVASIN--EGMTKRFSCITPODGQELVDGLKVCLOAALRAMNSCREM 607
 618 KSTRDKRAVAGALIASMDLQONSTYFTVTECSAFDVANLTMIAKALROYHERKL 677
 608 PSRIIVRBDVGDDGQOKTLTVNVEVPOFLDCKSIGRGY-----NPLTVIVKRVNTR 661
 678 PSRIIVRBDVGSSGLKOLFEEVVDIIEKKT---EVARVGLSPQALATVIVTRSMNTR 734
 662 FFAOAGGLQNLPGTVIDVEVTRPEWTDFTIVSGAVSSVSPHHVIVIDNSGLKPDH 721
 735 FFLNG-----QNEPPTIIVDVITLPERYDFYIVSQVQGVTSPTSYNVLISNGLSBEK 790
 722 IORLVYKCHITVYVNPVAVPAPCOYAHKLAFLVQGSIHREBNLSNRLYYL 775
 791 MOKLYYKCHLYVNSGTTVRVAVCOYAKALATVIGTMLHSIPQALKEKFFYL 844

RESULT 12

ABB62084
 ID ABB62084 standard; protein; 843 AA.

XX ABB62084;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13044.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li FMD, Myers EW,

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06187.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX Disclosure; SEQ ID NO 13044; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL16175), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB557737-AB572072). The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 843 AA;

Query Match 34.8%; Score 1428.5; DB 4; Length 843;

Best Local Similarity 38.0%; Pred. No. 1.4e-129; Mismatches 294; Conservative 163; Mismatches 284; Indels 33; Gaps 14;

17 SKTSSGIIYRLSTNHRFLTSRPMALYOHIDYMLMEARLRSLFQHEDLIGKCHA 76
 88 SKKGTGVPVWLQTNFRLKTPKEMRIYHYHFEPSIENPRVGVLSNHNLLGSGYL 147
 77 FDGILFLPKLQOKYTEVFSKTRNGEDVRIITITLNLPTSPCLQFYNIIFRLIKI 136
 148 FDGLQFLTRKFEQETIVLSGSKLDIEYKISIKFVGFISCAEPFLQVNLILRRSMKG 207
 137 KNLQOIGRNYVNPNDIPSHRLVYVPGFTSIIQYENSIMLCYDVSFKVLRSETVDF 196
 208 LNLVLGRNLPDPRAKIEIRFPMELMPYETSIRQHEKDIILGTETIKVMTETIYI 267
 197 MENFYQTEBHKQEOVSKEILGLVYLTKNKTYRVDIDWDQNPSTFKKADGSEVSF 256
 268 MRCSNPARH--QDEVKRVNVDLIVLTIDNNRTYRINDVDFQGTFRSTF--SCKGRDISF 324
 257 LEYRKQYNOEITDLKQPVIVSOPKRRRGGTLPEPAMLIPELCYITGIDTMRDNFV 316
 325 VEYLTQYINRIRHNPPLISK--NRDKALKTNASELVAVIPELCRTGLTAMRNENFOL 383

```

QY 317 MCDLAHTRLTPRQREVGRLIDYHKNDNVQ---RELADMGSLSPDNLSSFGRILOT 373
D 384 MRAMSSYTRMNPQR---TDRLAFNHRLOTPESVAVLDDNMNMLDKNTEVQGRITIGQ 440
QY 374 EKI--HOGGKTFDYNPQFADMSKETRGAPLISVYP--LDNMLLYTRRYEANSILQNL 429
D 441 QNIVFNHGKVPAGEN---ADMQRHFRDQRMULTPSDGLDRRAVIAPOGRNSHELRITLDSL 497
QY 430 FKVTAMGMQMRK-ALMIEVDRTETAYLRVLOQKVTADTQIVLCLSSNKDKYDAIKKY 488
D 498 YRAASGMGLRIRSRQEFIIYDDRGTGYVRAMDVCVRSBPKLITCLVENDNAERISSTIKR 557
QY 489 LCTDCTPSPCCVAVARTGKQOTVMAITKIALQNMCKMGELMRVDIPLKLVMIIVGIDCY 548
D 558 GYVDRAVPTQVITLKT-KRSLMSIATKIALQNLGTYPMIIEPLSGMTIGFDIA 616
QY 549 HDMTAGRSIAGFVASIN-EGMTWPSRCIFQDRGQELVDGLKVCLOALRANSCNRYM 607
D 617 KSTRDKRAVAGALIASMDLQONSTYFSTVTECSAFDVLANTLPMIAKALROYOHEHRKL 676
QY 608 PSRIIVRDGVGDQGLKTLVNYEVPQFLDCIKSIGRGY-----NPRLTIVYVKKRVNTR 661
D 677 PSRIIVRDGVSSGLQFLFEFVKDIEKLT---EYARVQSPPLATIVVTRSWNTR 733
QY 662 FFAOSGRLQNPFGTIVIDEVTRPEWYDFPIVSQAVRSGSVSPTHNVIVDSGLKPDH 721
D 734 FFLNG---QNPPTIIVDVITLPERYDFYLSQVQGVTSPTSYVNLVSSMGSLPEK 789
QY 722 IORLTYKCHLYNMPGVIRVPACQYAKLAFLVQGSIHREPULISNRLYYL 775
D 790 MOKLTYKCHLYNMSGTTRVPACQYAKLAFLVGNLHSHIPQNALKKFYLL 843

RESULT 13
AA90233
ID AA90233 standard, protein; 843 AA.
XX
AC AA90233;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila piwi protein.
XX
KW piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
XX anaemia; immunodeficiency; male infertility; Drosophila.
XX
OS Drosophila sp.
XX
FH Key location/Qualifiers
FT MISC-difference 13 /label= Leu, Ile
FT MISC-difference 106 /note= "encoded by NTT"
FT MISC-difference 106 /label= Leu, Ile
FT MISC-difference 785 /note= "encoded by NTA"
FT MISC-difference 785 /label= Leu, Ile
FT MISC-difference 785 /note= "encoded by NTC"
XX
XX WO200032039-A1.
XX
XX PD 08-JUN-2000.
XX
XX PF 03-DEC-1999; 99MO-US028764.
XX
XX PR 04-DEC-1998; 98US-0110901P.
XX
XX PA (UYDU-) UNIT DUKE.
XX
XX PI Lin H.
XX
XX WPI; 2000-412085/35.
XX
XX N-PSDB; AAA07586.
DR

```

```

XX
PT piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and diagnostic
PT applications.
XX
PS Claim 4, Page 171-175, 201pp; English.
XX
CC This sequence represents the Drosophila piwi family protein, designated
CC piwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat tissue
CC dystrophy, anaemia, immunodeficiency, and male infertility
XX
SQ Sequence 843 AA;

```

```

Query Match 34.5%; Score 1416.5; DB 3; Length 843;
Best Local Similarity 37.7%; Pred. No. 26-128;
Matches 292; Conservative 163; Mismatches 286; Indels 33; Gaps 14;

```

```

QY 17 SKTSSGIIVRLSTNHFRLTSRPPQALYOHIDYPMARLRBALPQHEDLIGKCHA 76
D 88 SKKGTGVPMQLQTFEFAKTRBEWRIYVHVEFPTLENPRGVGNHNLASGYL 147
QY 77 FQGTILFLPKRLOQKTEVSKTRNGEDVRITITLNELPPTSPCLQFYNIIFRLTKI 136
D 148 FDLQLFTTRKEQELTVLSGSKDIEKISIKFVGFISCAEPRLQVLTILRSKMG 207
QY 137 MNLQIGRNYNENDIDIPSHRLVIMPGFTTSILOYSIMLCTDVSHKTLRSEVLD 196
D 208 LNELVGRNLPFRAKISIREFEMELMPGETSIRHEKDILGHEITRKVMTETIYI 267
QY 197 MNPFGTEBHEKQEOVSKELGLVYLYTKYNNKTVVDDIDMDQNKSTFKADGSEVSE 256
D 268 MRCSNPAH--QDEVRYNVDLIVLTIDYNNRTYINDVDRCQTPKSTF-SCGRDISP 324
QY 257 LEYRKQNOETIDLKQPVLSQPKRRRGPGTLPPAMLIPELCYLTGLTDMRDNV 316
D 325 VEYLLTKNIRIRDNHQPILISK-NRDKALKTVAELVILPELCVGTGIMEMRNEFOL 383
QY 317 MCDLAHTRLTPRQREVGRLIDYHKNDNVQ---RELADMGSLSPDNLSSFGRILOT 373
D 384 MRAMSSYTRMNPQR---TDRLAFNHRLOTPESVAVLDDNMNMLDKNTEVQGRITIGQ 440
QY 374 EKI--HOGGKTFDYNPQFADMSKETRGAPLISVYP--LDNMLLYTRRYEANSILQNL 429
D 441 QNIVFNHGKVPAGEN---ADMQRHFRDQRMULTPSDGLDRRAVIAPOGRNSHELRITLDSL 497
QY 430 FKVTAMGMQMRK-ALMIEVDRTETAYLRVLOQKVTADTQIVLCLSSNKDKYDAIKKY 488
D 498 YRAASGMGLRIRSRQEFIIYDDRGTGYVRAMDVCVRSBPKLITCLVENDNAERISSTIKR 557
QY 489 LCTDCTPSPCCVAVARTGKQOTVMAITKIALQNMCKMGELMRVDIPLKLVMIIVGIDCY 548
D 558 GYVDRAVPTQVITLKT-KRSLMSIATKIALQNLGTYPMIIEPLSGMTIGFDIA 616
QY 549 HDMTAGRSIAGFVASIN-EGMTWPSRCIFQDRGQELVDGLKVCLOALRANSCNRYM 607
D 617 KSTRDKRAVAGALIASMDLQONSTYFSTVTECSAFDVLANTLPMIAKALROYOHEHRKL 676
QY 608 PSRIIVRDGVGDQGLKTLVNYEVPQFLDCIKSIGRGY-----NPRLTIVYVKKRVNTR 661
D 677 PSRIIVRDGVSSGLQFLFEFVKDIEKLT---EYARVQSPPLATIVVTRSWNTR 733
QY 662 FFAOSGRLQNPFGTIVIDEVTRPEWYDFPIVSQAVRSGSVSPTHNVIVDSGLKPDH 721
D 734 FFLNG---QNPPTIIVDVITLPERYDFYLSQVQGVTSPTSYVNLVSSMGSLPEK 789
QY 722 IORLTYKCHLYNMPGVIRVPACQYAKLAFLVQGSIHREPULISNRLYYL 775
D 790 MOKLTYKCHLYNMSGTTRVPACQYAKLAFLVGNLHSHIPQNALKKFYLL 843

```

RESULT 14

Wed May 12 09:50:21 2004

us-10-043-774b-2.rag

Page 13

activity of polynucleotides of the invention may be described as:
CC vulnerable, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABP43544-
CC ABP43989 represent polypeptides encoded by polynucleotides of the
CC invention. NOTE: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pct_sequences

XX Sequence 498 AA:

Query Match 32.3%; Score 1324.5; DB 5; Length 498;

Best Local Similarity 51.4%; Pred. No. 8.3e-120; Indels 9; Gaps 4;

Matches 259; Conservative 86; Mismatches 150;

QY 275 VLSQPKRRGPGGTLPGPMLIPELCYLTGLDKKNDNFMKDLAVHRLPTEQORE 334
DB 1 MLVSLKKKKNDSE-PQLAHLIPELCYLTGLDQATSDPOLKAVAKRSLSPSGRQR 59
QY 335 VGRLLDYHKKNDVQRELDMGSLFSDNLSFGRLIQTEKHOGGKTFDYNPQ---PAD 391
DB 60 LARLVNIOKNTARFELTNGLHFGSQ-ISLGRIVPSEKILMQ---DHICQVSAAD 114
QY 392 WSKETRAGAPLISVPLDNMLIYTRYEANSILQNLFKYTPAMGQMKALMIEVDR 451
DB 115 WSKDIRCKILMAQSLMTWLICSDRTEYVAESFLNCLRRVAGSWGPNVDYPKIKQEN 174
QY 452 TEALVAVQCKVTADTOIVVCLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGQQT 511
DB 175 PAATVRAIQYVDPDQVLMCILPSNOKTYDSIKKYLSDCPVPSQCVLARTLKGMM 234
QY 512 MAITKIALQNNCKNGELWEVDIPLKLVMTVIGIDCHDMTAGRSIAGFASINEMTR 571
DB 235 MSIAFKIMQMTCKLGELMAVEIPLKSLMVGIDVCKDALSQVMVVGCVASVNERITR 294
QY 572 WFSRCIFQDRQGLVDGLKVCLOALRAMNSCNEYMPRIIVYRDGVGDGQLKTLVNYEV 631
DB 295 WFSRCILQRTWTDVADCLKVFETGALNKRYKYNHDLPARIIYRAGVGDGQLKTLIEYEV 354
QY 632 PQFLDCKSIGRGNPRLTVIVKKRVNTRFFAQSGLQNPDPETVIDVEVTRPEWTF 691
DB 355 PQLSSVASSSNTSRLSVIIVRAKKCMRFTENKRTVONPLGTIVDSSEATRENEWDF 414
QY 692 FIVSQAVRSGSVPTHTNVIYNSGLKPDHIOQLTYKCHITYNMPGIVRVPAPQYAHK 751
DB 415 YLISQVACRGTVSPYYNVIYDNGIKPDHMQRLTKLCHLYNMPGIYSVAPPCQYAHK 474
QY 752 LAFIVGQSIHREPNSLSNRLYYL 775
DB 475 LTFLGQSIHKPSLELANHLFYL 498

Search completed: May 5, 2004, 15:12:38
Job time : 63 secs

Wed May 12 09:50:22 2004

us-10-043-774b-2.rsp

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:06:35 ; Search time 17 Seconds
(without alignments)

2373.786 Million cell updates/sec

Title: US-10-043-774B-2

Sequence: 1 MIFGVNTRQNDHVKESKGTG.....VQSIRHBPNTLSNRLYYL 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1428.5	34.8	843	1	PIWI DROME
2	557.5	13.6	834	1	IF2C SCPO
3	510	12.4	1048	1	AGOI ARATH
4	508.5	12.4	860	1	I2C3 MOUSE
5	506.5	12.3	857	1	I2C1 HUMAN
6	504.5	12.3	860	1	I2C2 RAT
7	504.5	12.3	861	1	I2C4 MOUSE
8	503.5	12.3	840	1	I2C2 RABIT
9	503	12.3	988	1	PIWI ARATH
10	502	12.2	860	1	I2C3 MOUSE
11	499.5	12.2	857	1	I2C1 MOUSE
12	497	12.1	860	1	I2C3 HUMAN
13	495.5	12.1	861	1	I2C4 HUMAN
14	495.5	12.1	997	1	AGOI ARATH
15	487	11.9	851	1	I2C2 HUMAN
16	482	10.3	1040	1	I043 CAEL
17	280.5	6.8	945	1	GCC7 CAEL
18	248.5	6.1	1032	1	Y053 CAEL
19	145	3.5	773	1	YD21 METU
20	137.5	3.4	1912	1	DICE HUMAN
21	134.5	3.3	1906	1	DICE MOUSE
22	118	2.9	1120	1	MED RICEP
23	114	2.8	2035	1	HMP2 YEREN
24	113.5	2.8	560	1	PHAC PSEOL
25	113	2.8	949	1	PM46 ARATH
26	113	2.8	3144	1	VP13 YEAST
27	111.5	2.7	1909	1	DICE ARATH
28	108.5	2.6	1369	1	TRAP HUMAN
29	108	2.6	1317	1	N145 YEAST
30	108	2.6	1653	1	CIH YEAST
31	107	2.6	1282	1	SN3X MOUSE
32	106.5	2.6	690	1	Y173 URPPA
33	104.5	2.5	1966	1	DOC4 HUMAN

34	104	2.5	1180	1	C4AA BACTI	P16480 bacillus th
35	104	2.5	1277	1	PDS5 YEAST	O04264 saccharomyc
36	104	2.5	4128	1	PRKD MOUSE	P97313 mus musculu
37	103	2.5	1342	1	RPOB HAEIN	P43738 haemophilus
38	103	2.5	1575	1	I062 HUMAN	O13576 homo sapien
39	102.5	2.5	1575	1	CHLI CHLVU	P26304 chlorella v
40	102.5	2.5	869	1	YD95 YEAST	P38966 saccharomyc
41	102.5	2.5	1170	1	IT2A BOVIN	P53710 bos taurus
42	102	2.5	3122	1	DEOZ MOUSE	O61493 mus musculu
43	101.5	2.5	3135	1	S230 PLAF0	O08372 plasmodium
44	101	2.5	1388	1	RPOD TOBAC	P38550 nicotiana t
45	101	2.5	1886	1	POL COYIV	P19199 commelina y

ALIGNMENTS

RESULT 1
PIWI DROME STANDARD; PRT; 843 AA.
ID PIWI DROME
AC O9VXMI, 096674; 096675;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PIWI protein
GN PIWI OR CG6122.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
[2]
SEQUENCE FROM N.A., FUNCTION, DEVELOPMENTAL STAGE, AND TISSUE
SPECIFICITY.
RC STRAIN=Canton-S, and Oregon-R.
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.,
RT "A novel class of evolutionarily conserved genes defined by piwi are
RT essential for stem cell self-renewal."
RL Genes Dev. 12:3715-3727(1998).
[2]
SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amaralides P.G., Scherer S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Burtis K.C., Busam D.A., Butler H., Caden E., Center A., Chandra I.,
RA Burris K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Chertys J.M., Cullen P., Deng Z., Mayes A.D., Dew I., Dietz S.M.,
RA de Pablos B., Delcher A., Deng Z., Mayes A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D.A., Heiman T.U., Hernandez J.R., Houck J.,
RA Hovstlin D., Houston K.A., Howland T.U., Wei M.-H., Idegawa C.,
RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laaso P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mikhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacheco J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,


```

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huxst S.M.,
RA Lucas M., Rocher M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Ruenzel U.J., Moreno S., Armstrong J., Forsberg S.L.,
RA Cerutti L., Lowe T., McComble W.R., Palsten I., Potashkin J.,
RA Shpakovski G.V., Usery D., Bartell B.G., Nurre P.;
RT "The genome sequence of Schizosaccharomyces pombe." ;
RL Nature 415:871-880(2002).
CC -1- FUNCTION: plays an important role in the eukaryotic peptide chain
CC initiation process (By similarity).
CC -1- SIMILARITY: Belongs to the argonate family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 P1wt1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AL023705; CAA19275.1; -
CC PIR; T41568; T41568.
CC GeneDB SPombe: SPCC736.11; -.
DR GO; GO:00005717; C:chromatin; ISS.
DR GO; GO:00005694; C:chromosome; ISS.
DR GO; GO:0016263; C:eukaryotic 48S initiation complex; ISS.
DR GO; GO:0005654; C:nucleoplasm; ISS.
DR GO; GO:0005718; C:nucleosome; ISS.
DR GO; GO:0007049; P:cell cycle; ISS.
DR GO; GO:0010043; P:cell organization and biogenesis; ISS.
DR GO; GO:0006335; P:establishment and/or maintenance of chromatin; ISS.
DR GO; GO:0006412; P:protein biosynthesis; ISS.
DR GO; GO:0006355; P:regulation of transcription; ISS.
DR GO; GO:0006350; P:transcription; ISS.
DR InterPro; IPR003160; PAZ.
DR InterPro; IPR003165; P1wt1.
DR Pfam; PFO2171; P1wt1.
DR Pfam; PFO2170; PAZ.
DR PROSITE; PS02171; P1wt1.
DR PROSITE; PS08822; P1wt1.
DR KEGG; K01001; P1wt1.
DR KEGG; K01002; PAZ.
DR KEGG; K01003; P1wt1.
DR KEGG; K01004; PAZ.
DR KEGG; K01005; P1wt1.
DR KEGG; K01006; PAZ.
DR KEGG; K01007; P1wt1.
DR KEGG; K01008; PAZ.
DR KEGG; K01009; P1wt1.
DR KEGG; K01010; PAZ.
DR KEGG; K01011; P1wt1.
DR KEGG; K01012; PAZ.
DR KEGG; K01013; P1wt1.
DR KEGG; K01014; PAZ.
DR KEGG; K01015; P1wt1.
DR KEGG; K01016; PAZ.
DR KEGG; K01017; P1wt1.
DR KEGG; K01018; PAZ.
DR KEGG; K01019; P1wt1.
DR KEGG; K01020; PAZ.
DR KEGG; K01021; P1wt1.
DR KEGG; K01022; PAZ.
DR KEGG; K01023; P1wt1.
DR KEGG; K01024; PAZ.
DR KEGG; K01025; P1wt1.
DR KEGG; K01026; PAZ.
DR KEGG; K01027; P1wt1.
DR KEGG; K01028; PAZ.
DR KEGG; K01029; P1wt1.
DR KEGG; K01030; PAZ.
DR KEGG; K01031; P1wt1.
DR KEGG; K01032; PAZ.
DR KEGG; K01033; P1wt1.
DR KEGG; K01034; PAZ.
DR KEGG; K01035; P1wt1.
DR KEGG; K01036; PAZ.
DR KEGG; K01037; P1wt1.
DR KEGG; K01038; PAZ.
DR KEGG; K01039; P1wt1.
DR KEGG; K01040; PAZ.
DR KEGG; K01041; P1wt1.
DR KEGG; K01042; PAZ.
DR KEGG; K01043; P1wt1.
DR KEGG; K01044; PAZ.
DR KEGG; K01045; P1wt1.
DR KEGG; K01046; PAZ.
DR KEGG; K01047; P1wt1.
DR KEGG; K01048; PAZ.
DR KEGG; K01049; P1wt1.
DR KEGG; K01050; PAZ.
DR KEGG; K01051; P1wt1.
DR KEGG; K01052; PAZ.
DR KEGG; K01053; P1wt1.
DR KEGG; K01054; PAZ.
DR KEGG; K01055; P1wt1.
DR KEGG; K01056; PAZ.
DR KEGG; K01057; P1wt1.
DR KEGG; K01058; PAZ.
DR KEGG; K01059; P1wt1.
DR KEGG; K01060; PAZ.
DR KEGG; K01061; P1wt1.
DR KEGG; K01062; PAZ.
DR KEGG; K01063; P1wt1.
DR KEGG; K01064; PAZ.
DR KEGG; K01065; P1wt1.
DR KEGG; K01066; PAZ.
DR KEGG; K01067; P1wt1.
DR KEGG; K01068; PAZ.
DR KEGG; K01069; P1wt1.
DR KEGG; K01070; PAZ.
DR KEGG; K01071; P1wt1.
DR KEGG; K01072; PAZ.
DR KEGG; K01073; P1wt1.
DR KEGG; K01074; PAZ.
DR KEGG; K01075; P1wt1.
DR KEGG; K01076; PAZ.
DR KEGG; K01077; P1wt1.
DR KEGG; K01078; PAZ.
DR KEGG; K01079; P1wt1.
DR KEGG; K01080; PAZ.
DR KEGG; K01081; P1wt1.
DR KEGG; K01082; PAZ.
DR KEGG; K01083; P1wt1.
DR KEGG; K01084; PAZ.
DR KEGG; K01085; P1wt1.
DR KEGG; K01086; PAZ.
DR KEGG; K01087; P1wt1.
DR KEGG; K01088; PAZ.
DR KEGG; K01089; P1wt1.
DR KEGG; K01090; PAZ.
DR KEGG; K01091; P1wt1.
DR KEGG; K01092; PAZ.
DR KEGG; K01093; P1wt1.
DR KEGG; K01094; PAZ.
DR KEGG; K01095; P1wt1.
DR KEGG; K01096; PAZ.
DR KEGG; K01097; P1wt1.
DR KEGG; K01098; PAZ.
DR KEGG; K01099; P1wt1.
DR KEGG; K01100; PAZ.
DR KEGG; K01101; P1wt1.
DR KEGG; K01102; PAZ.
DR KEGG; K01103; P1wt1.
DR KEGG; K01104; PAZ.
DR KEGG; K01105; P1wt1.
DR KEGG; K01106; PAZ.
DR KEGG; K01107; P1wt1.
DR KEGG; K01108; PAZ.
DR KEGG; K01109; P1wt1.
DR KEGG; K01110; PAZ.
DR KEGG; K01111; P1wt1.
DR KEGG; K01112; PAZ.
DR KEGG; K01113; P1wt1.
DR KEGG; K01114; PAZ.
DR KEGG; K01115; P1wt1.
DR KEGG; K01116; PAZ.
DR KEGG; K01117; P1wt1.
DR KEGG; K01118; PAZ.
DR KEGG; K01119; P1wt1.
DR KEGG; K01120; PAZ.
DR KEGG; K01121; P1wt1.
DR KEGG; K01122; PAZ.
DR KEGG; K01123; P1wt1.
DR KEGG; K01124; PAZ.
DR KEGG; K01125; P1wt1.
DR KEGG; K01126; PAZ.
DR KEGG; K01127; P1wt1.
DR KEGG; K01128; PAZ.
DR KEGG; K01129; P1wt1.
DR KEGG; K01130; PAZ.
DR KEGG; K01131; P1wt1.
DR KEGG; K01132; PAZ.
DR KEGG; K01133; P1wt1.
DR KEGG; K01134; PAZ.
DR KEGG; K01135; P1wt1.
DR KEGG; K01136; PAZ.
DR KEGG; K01137; P1wt1.
DR KEGG; K01138; PAZ.
DR KEGG; K01139; P1wt1.
DR KEGG; K01140; PAZ.
DR KEGG; K01141; P1wt1.
DR KEGG; K01142; PAZ.
DR KEGG; K01143; P1wt1.
DR KEGG; K01144; PAZ.
DR KEGG; K01145; P1wt1.
DR KEGG; K01146; PAZ.
DR KEGG; K01147; P1wt1.
DR KEGG; K01148; PAZ.
DR KEGG; K01149; P1wt1.
DR KEGG; K01150; PAZ.
DR KEGG; K01151; P1wt1.
DR KEGG; K01152; PAZ.
DR KEGG; K01153; P1wt1.
DR KEGG; K01154; PAZ.
DR KEGG; K01155; P1wt1.
DR KEGG; K01156; PAZ.
DR KEGG; K01157; P1wt1.
DR KEGG; K01158; PAZ.
DR KEGG; K01159; P1wt1.
DR KEGG; K01160; PAZ.
DR KEGG; K01161; P1wt1.
DR KEGG; K01162; PAZ.
DR KEGG; K01163; P1wt1.
DR KEGG; K01164; PAZ.
DR KEGG; K01165; P1wt1.
DR KEGG; K01166; PAZ.
DR KEGG; K01167; P1wt1.
DR KEGG; K01168; PAZ.
DR KEGG; K01169; P1wt1.
DR KEGG; K01170; PAZ.
DR KEGG; K01171; P1wt1.
DR KEGG; K01172; PAZ.
DR KEGG; K01173; P1wt1.
DR KEGG; K01174; PAZ.
DR KEGG; K01175; P1wt1.
DR KEGG; K01176; PAZ.
DR KEGG; K01177; P1wt1.
DR KEGG; K01178; PAZ.
DR KEGG; K01179; P1wt1.
DR KEGG; K01180; PAZ.
DR KEGG; K01181; P1wt1.
DR KEGG; K01182; PAZ.
DR KEGG; K01183; P1wt1.
DR KEGG; K01184; PAZ.
DR KEGG; K01185; P1wt1.
DR KEGG; K01186; PAZ.
DR KEGG; K01187; P1wt1.
DR KEGG; K01188; PAZ.
DR KEGG; K01189; P1wt1.
DR KEGG; K01190; PAZ.
DR KEGG; K01191; P1wt1.
DR KEGG; K01
```

Db	288	K1SVAEYLENNNRVLRQVBN.PCIL-----VNKGAMLP1EFCFVVGKGRYAK	333
Qy	310	MENDENVMKDLAVHTRLTTPROQROBEVGLDHYHKAD-VNORELRDMGSLSPDNLISFSG	368
Db	334	LMSD-----QTNMTRFAVORPFRFVQOIDDVFVQMDMDDDPYLTQGMKIQQKMLEVPA	388
Qy	369	RLIOTEKIHOGAKTVDYVQPADMSKETRGALISV--KEPLDWLLI---YTRR-NYEA	422
Db	389	RYLETPSIRYSGDCLIE-RPVSGRNN--LRGRFLDPPRAPLRISAVVACFSTRLRPMGI	445
Qy	423	NSLIQNLFFKVTBPMGMQ--MKRAIMIEVDN--TEAYLRVLQCKV-----TADTQIVVCLL	474
Db	446	ENPLQYVQVQTLTSLGINFVWKKPRLVYADIGSVEBELCITLYKKAEQVGNAPPDYLFLL	505
Qy	475	SSNRKDKDAIKKYLCTDPCPTPSQCVAVARLTKQQTWALATKALQMKMGG---ELW	531
Db	506	DKNSPEPYSIRVCNTMLGVPSQCALSKH1--LQSKPYQCANLGMKINRVKGINCSLI	563
Qy	532	RVDP1EKLTV--MIVGIDCYH--DMTAGRSIAGFVASINEGTWFMSRCIFQDRGQELVDG	588
Db	564	PKSNPLGKVPFLILGDVYHHEVGATGVSIASIVADLNGCKTAVRSQCPHQEVIIEG	623
Qy	589	LKYCTGALPAMNSCNEIMPSEIIYTRDGVGDQQLTVLVEYVQFLDCLKSLKRGVNR	648
Db	624	MKQIVVYL1QGGFRATKQOPOKILITFRDGTSEGGFLSVINDELSQIEACHSISPKINPK	683
Qy	649	LTVIVVKKRVMTRFPAOS---CGRLQNP1PFTV1DVEVTRPEWDFEIVSQAVRSQSVSP	705
Db	684	ILVCTTQKHHRRFFIKKNSDDGRGNLPGTIIIEKVTHHYQDVFILSHPSLQGVSVF	743
Qy	706	THYVTVYDNSGKRPDHIOQLTYK1CHIYVMPGVIRVPAPCQVAKHAKFL-----	755
Db	744	VHYTVLHDEIQPBPQFLCYNLCTCYARATASVSLVPVYVHALVSNLARVQDVTADD	803
Qy	756	-----VGQSIHREP1LS1NLT 772	
Db	804	TFVETSEASMDQEV--KPLLALUSKLT 827	
RESULT 3			
AGOL_ARATH			
ID	AGOL_ARATH	STANDARD;	PRT; 1048 AA.
AC	004379; Q9LP83;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Argonaute protein.		
GN	AGOL1 OR AT1G48410 OR FL1A17.3 OR T1N15.2.		
OS	Arabidopsis thaliana (Mouse-ear cress); Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Streptophyta; eudicotyledons; core eudicots; rosids;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxId=3702;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv Columbia; TISSUE=leaf;		
RC	MEDLINE=21016719; PubMed=11130712;		
RA	Medlogis A., Becker U.R., Palm C.U., Federspiel N.A., Kaul S.,		
RA	White O., Alonso J., Atalfi H., Araujo R., Bowman C.L., Brooks S.Y.,		
RA	Buhler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,		
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,		
RA	Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,		
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzar L.,		
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.		
RA	Kim C.J., Koo H., Kremenetskaia I., Kuttz D.B., Kwan A., Lam B.,		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv Columbia;		
RC	MEDLINE=21016719; PubMed=11130712;		
RA	Medlogis A., Becker U.R., Palm C.U., Federspiel N.A., Kaul S.,		
RA	White O., Alonso J., Atalfi H., Araujo R., Bowman C.L., Brooks S.Y.,		
RA	Buhler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,		
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,		
RA	Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,		
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzar L.,		
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.		
RA	Kim C.J., Koo H., Kremenetskaia I., Kuttz D.B., Kwan A., Lam B.,		

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maltz R., Marshall A.,
 RA Maltzsch J., Miranda M., Nguyen M., Niemann W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
 RA Sun H., Tallon L.V., Tambunga G., Tortum M.J., Town C.D.,
 RA Uteback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Tortum M.J.,
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Anarri Y.,
 RA Chakawa T., Bani N., Banno F., Bowser L., Brooks S.Y., Cathinici P.,
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
 RA Kaniya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
 RA Saito M., Tanase R., Vaysberg M., Walender E.K., Wong C., Yamamura Y.,
 RA Yano S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
 RT "Empirical analysis of transcriptional activity in the Arabidopsis
 genome.";
 RL Science 302:842-846(2003).
 CC -1- FUNCTION: Essential for proper development of leaves and floral
 CC organs and formation of axillary meristems.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=A number of isoforms are produced. According to EST
 CC sequences;
 CC Name=1;
 CC IsoId=004379-1; Sequence=displayed;
 CC TISSUE SPECIFICITY: Widely expressed at low levels.
 CC -1- DEVELOPMENTAL STAGE: Expressed throughout all developmental
 CC stages.
 CC -1- SIMILARITY: Belongs to the argonaute family.
 CC -1- SIMILARITY: Contains 1 PAZ domain.
 CC -1- SIMILARITY: Contains 1 PIWI domain.
 CC -1- CAUTION: Ref.2 (AA79718) sequence differs from that shown due to
 CC erroneous gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC DR EMBL, U91995; AAC18440.1; -;
 CC DR EMBL, AC007932; AAD49755.1; -;
 CC DR EMBL, AC020889; AAF79718.1; ALT_SEQ.
 CC DR EMBL, BT000941; AAN4341.1; -;
 CC DR InterPro; IPR003100; PAZ.
 CC DR InterPro; IPR003165; PIWI.
 CC DR Pfam; PF02170; PAZ; 1.
 CC DR Pfam; PF02171; PIWI; 1.
 CC DR PROSITE; PS50821; PAZ; 1.
 CC DR PROSITE; PS50822; PIWI; 1.
 CC DR Developmental protein; Alternative splicing.
 CC FT DOMAIN 391 501 PAZ.
 CC FT DOMAIN 676 997 PIWI.
 CC FT DOMAIN 13 104 GLY-RICH.
 CC SEQUENCE 1048 AA; 116190 MW; 3E514633A09C41 CRC64;
 CC Query Match 12.4%; Score 510; DB 1; Length 1048;

Best Local Similarity 25.1%; Pred. No. 9,2e-29;
 Matches 213; Conservations 136; Mismatches 376; Indels 124; Gaps 33;
 QY 20 GSSGIIVRLSTNHRFLRSPPQWALYYHIDYNPIMEARLRSLFPQEDLIGCH--- 75
 DB 186 GSGKRCIVKANHF-FAELPDKDLHYDVITTEVTSRGVAVMKQLVNYSRSHGSR 244
 QY 76 -ADG-----TILFL-----KLOOKTEVSKTRNGEDVRI 109
 DB 245 LPAYDKRSKYTAGPLPPNSKEFFINLLDEVGAGGRRRRFRVIVKYVARDLHLGM 304
 QY 110 TLTELPPTSPCTQFYNIIFRLLKIMLQOIGRNYNPNDPIDPSH---LVYWP 164
 DB 305 FLEKQSDAPQEAQVADIVLRE-LPTSRYPVGRSFYSP---DIGKQSLDGLSWR 359
 QY 165 GFTTSILOEYSINLCTDVSHKV-LREPTVLDENFNYHOTBEK-----FOEVSKELI 218
 DB 360 GFYOSIRPTQWGLSINIDMSSTAFIEANPVQFCDLLNRDISRPISDADRVYIKALR 419
 QY 219 GL-VLTQKN--KTYRVDIDMDQNPSTFKAD-GSEVSFLRYEKYNOEITDKOP 274
 DB 420 GVKVEYTRGMRKRYRISGLTAVATBELFPVDERNTQGSVEYFHEYGFRIQHTOLP 479
 QY 275 VLVSPKRRRPGGTLPPAPMLPELCYLT--GLTDKRDENVMKDLAVHTRILPBRQ 332
 DB 480 CLQVNSNR-----PNYLPMEVCKIVGQRYSKRLRQITALLKV-TQRPIDRE 529
 QY 333 REV---GRLLDYHNDVQSELDDWGLSPSNLSFSGRLQTE--KHGSGKTFDYNP 387
 DB 530 KQILQTVQNDY--AKDYVADE--FGKISTLSLVAZARLLPPMKYHSGREGTCLP 584
 QY 388 QFADMSKETRGAPLISVPLDNWLLIYRNYE--AANSLQNLKVTTPANGQMR-KAI 444
 DB 585 QVGQNNMNR--KMINGTVANNWICINFSRGVODNLATFOELQAWQYVSGMAFNPBPV 642
 QY 445 MEVDRTREAVIRLVQCKVTADT-----QIVCLLSNRKDKYDAIKKYCTDPT 495
 DB 643 LPVSAREQVEKVLKTYHDATSKLSQGEIDLLIVLPNNNGSLYDDLRIGETELGI 702
 QY 496 PQCVAVARTLQ-QQVNAIAIKIALQWCKNGG-ELMRVD---TL--KVMYIGI 545
 DB 703 VSQCLTGVFPMRSQVNA--NVALKINVKYGGHNTLVVALSRILVSDRPTIRGA 759
 QY 546 DCTHMTA--GRSLIAGVYASIN-EGMTRWTSRCLFQDROGELVDG----- 589
 DB 760 DVTTHHPGDSFSLAAVVASQDMPETKIGLVCAQHRQELIDDLKEMQDPKGYVT 819
 QY 820 GGMTEKILLAFRSTGKPLRIIFRDVSESGFQVLLYELDIAIRKACASLEAGYOPPV 879
 QY 650 TIVVKKRNTFFPQ-----SGRLNPPLPGTYIDVETPEPEVYFVVSQAVRSGSV 703
 DB 880 TVVVQKRRHTLFPQNHNDHSVDRSGNIDPGLVDSKICPTPEFDYLSHAIGQIS 939
 QY 704 SPTHVIVYDNGSLKPDHIOQLTYKLCIYNNPQVIRVAPACQYAHKIAFLVGSIHRE 763
 DB 940 RPAHVHVLNDENNFTADGLQSLTNMLCYTACRSVSVIPPAYVAHAAFPRA--RYWE 997
 QY 764 PNLISNRL 772
 DB 998 PETSDSGM 1006
 RESULT 4
 ID12C2_MOUSE STANDARD; PRT; 860 AA.
 AC Q8CUG0;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
 DE (P1WI/argonaute family protein meIP2C2).

```

GN EFPC2.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN NCBI_TaxId=10990;  
RP SEQUENCE FROM N.A.  
RA doi N., Zenko S., Ut-Tel K., Takahasi F., Ueda R., Miyata Y.,  
SA Satoh K.,  
RT "elf2C family proteins and Dicer homologs are required for siRNA  
mediated RNAi in mammalian cells";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
CC -! FUNCTION: Plays an important role in the eukaryotic peptide chain  
initiation process (By similarity).  
CC -! SIMILARITY: Belongs to the argonaute family.  
CC -! SIMILARITY: Contains 1 PAZ domain.  
CC -! SIMILARITY: Contains 1 Piwi domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation at  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
CC EMBL; AB081472; BAC1567.1; --  
DR MGd; MG1:244632; Elf2c2.  
DR DR InterPro; IPR003100; PAF.  
DR InterPro; IPR003165; PIWI.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; PIW1; 1.  
DR PROSITE; PS00821; PAZ; 1.  
DR PROSITE; PS00822; PIW1; 1.  
DK Initiation factor; Protein biosynthesis.  
FT DOMAIN  
      236     349  
            PAZ.  
FT FT  
      518  
            PIW1.  
SQ DOMAIN  
      860 AA: 97294 MW; DA32A759EA467B31 CRC64;
```

Query Match 12.4% Score 508.5 DB 1 Length 860;
Best Local Similarity 23.8%; Pred No. 8.ee-29;
Matches 195; Conservative 152; Mismatches 350; Indels 121; Gaps 32.

```
QY      20 GSSGIVLELTSTNHFELTSRPOVALYOHIIDYNPLMEARRLSAL--LPQH-BDLIG-K 73  
       1:::|::|:  
Db      33 GTTGRTIKLANFFEM-DIPKIDIHYELDIFKEGRPARVREIYEHWGHFKIQDIFEDR 91  
               |::||::|:  
QY      74 CHAPFG----TLLEPKLRLOQKVETVSQRNRNGD-VRLTI-----TLTN 113  
                |||::|::|:  
Db      92 KPVEFGRKNLYTMFLPTG-RDKALEVTLPGEGBORLIKSIRKVCVSIALDALDSG 150  
                 |||::|::|:  
QY     114 ELPEPSPCYLFONYNIIEFRLLRIKWMLQQIGHNYNPNDPIDP-SHRLVIWPGETSLIQ 172  
              ::||::|:  
Db     151 RLPSVFEPETLDALDVMTSH-LPSMKYTEVGSRSEFPASGCNSLPQGGREVWFGRHSVRP 209  
                  :||:  
QY     173 YENSIIMCTVDISHKL-RSETVLDP--MNFPHYOTESHK-----FOGVSKELLGLVYL 223  
           |||::|:  
Db     210 SLKKMMNLINIVSNAFYAQAPIEFCCELVDFKISIEOQKPITDSQRYKFTEKIKGAYE 269  
          ::||:  
QY     224 TKY---NNKTAVDDIDWDQNPFSTFKKADS---EVSFLEYRKQYNQEITLKQPVL 276  
         ::||::|:  
Db     270 ITHGCGMRKRRCVNVTRRRPASPHOFPLQGSGGTVECTVAQYFDHRKHVLAERHPCL 329  
        277 VSQPKRRRGPGGITLGPMILIPELCYTIG-----LTDGRNDEVMKDIAVHTRLTP 328  
             |||:  
Db     330 QVGEGEQK-----TYPLEVCNIVAQCRCIKTLTDQTSTM-----IRATARSAP 374  
            :||:  
QY     339 BOROREVARLIDYIHKNDNVORELBDWLSDPDSNLSFSGRILLOTEKIHOGAKTFDVPNQ 388  
          ::||::|:  
Db     375 D-ROEKESIKLNK--SASFNTDPYFAEFGINMYKDEMTDYAGRLDPSSLIIYGR----- 424  
                    :||:  
QY     389 PADWSKERGAPLLISVKELDN-----WLLIYTRYNEYAA--NSLIONTKYTV 433
```

Db	425	-----NKAIATPVGWMDMKKHQHTIELIKWALIAAFAPRQCTEVLHLSFTGLKRS	478
Qy	434	PAMGWQKPK-----AIMEVDPRFEAYLFIYIQQTATDTQIYVCLLSNRDKYAIKYL	489
Db	480	RDAMPMPLOGPBCFCFKYAGADSVSEPMFRHL-KNTYAGLQIYVIL--DGKIPVYAEVRKG	537
Qy	490	CTDGPPTSQCVAATLTKGQGTWALAKILQNNCKMG-----ELMRVIDPLKLVIV	543
Db	538	DTVIGMATQCYQMKNV--QRTPPTLSNLCKIKIVKLGYYNNILPQGRPPVFGQPIFL	595
Qy	544	GIDCYHMDTAG--RSIAGFVGSINSGMTRWFSRCIFDRGQELVGLKVCQLQALRAW	600
Db	596	GADVTHP-PADGKKPPIAAVVGSMADHPNRYCATVAVQOHQRIIODLAAMVRELLIQF	654
Qy	601	NSCHWNPESRIYVRDVGQGLKTYVNYVPEPLDCLKSIGRQYRNLVIYKKEVNT	660
Db	655	YKSTRFPKTRIFIFRDVSEQFOQVLAHMLAIRCEIKLEKDYQGTITFVIQKSHHT	714
Qy	661	RFPA-----QSGGRLQNLPLPGVILDEVTREBWDYFFIVSQAVBSGVSPTHYNYVNSG	716
Db	715	RLFTCTDNERNRGKSGNIPAGTYVDTKLTHTPEPFYICSHAGIGTGRPSHYHLWMDNR	774
Qy	717	LKPDHIQLRYKLCHTIYNNPQVIRPAPCQYAHKLAFL	754
Db	775	FSSDELQILTYQLCHTYVRCRHSVISPAPYVYHVLAVF	812
RESULT 5			
12CL	HUMAN		
ID	12CL HUMAN	STANDARD,	PRT; 857 AA.
AC	Q9ULI8;		
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1) (Putative RNA-binding protein Q99).		
GN	EIF2C1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=20005943; PubMed=10534406;		
RA	Koesters R., Adams V., Betts D., Moos R., Schmid M., Stermann A.,		
RA	Haasam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,		
RA	Briner J.;		
RT	"Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence, genomic organization, localization to chromosomal bands 1p34-p35, and expression."		
RT	Genomics 61:210-218 (1999).		
RL	-1- FUNCTION: Plays an important role in the eukaryotic peptide chain initiation process.		
CC	-1- SIMILARITY: Belongs to the argonaute family.		
CC	-1- SIMILARITY: Contains 1 PAZ domain.		
CC	-1- SIMILARITY: Contains 1 Pwi domain.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF093097; AAF00068.1; -.		
DR	GeneW; HGNC:3262; EIF2C1.		
DR	MIW; 606228; -.		
DR	GO; GO:0005050; C:eukaryotic translation initiation factor 2. . .; TAS.		
DR	GO; GO:0003743; F:translation initiation factor activity; TAS.		
DR	GO; GO:0006446; P:regulation of translational initiation; TAS.		
DR	InterPro; IPR003165; Pwi.		

PFam; PF02170; PAZ; 1.
 DR PFam; PF02171; P1W1; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; P1W1; 1.
 DR Initiation factor; Protein biosynthesis.
 FT DOMAIN 226 346
 FT 515 816 P1W1.
 SQ SEQUENCE 857 AA; 97214 MW; 1DBB524AE7CBAF66 CRC64;

Query Match 12.3%; Score 506.5; DB 1; Length 860;
 Best Local Similarity 24.1%; Pred. No. 1.7e-28;
 Matches 203; Conservative 160; Mismatches 349; Indels 133; Gaps 37;

20 GSSGIVRLSTNHFRLTSRPMALYOHYHIDYPLMEARRLSAL--LFQH--EDLIG-K 73
 30 GTVGKPIFLANFEV--DIPKIDYHYEVDIKPKDCPRVREVVEYVQHFKQIFGDR 88
 74 CHAFDQ-----TIFLPK--KELQKTVESKTRNGED--VRIT-----T 110
 89 KPYDGKKNIYVATLPNGRNVDFEYV---IPGSKDRIFKYSIKWLAIVSRMLHEA 144
 111 LTNELEPTSPCLQFYNIIFRLIKIMLQIGHYVNPNDPIDIP--SHRLVIMPGETTS 169
 145 LVSGQIPVPLESVQALDVAMRHLS--KRYTPVGRSFPSPREGYHPLGGREWFGRHS 203
 170 ILQYENSIMLCTDVSHKVL--RSEVLDPMF-----NFTQTEHKEQOY--SKELIG 220
 204 VRPAMMMKMLNIDVSATAFYKQPIVIEPMCEVLDIRNIDQPKLTDGQRFRTKEIKGL 263
 221 VILRYK--NKKTYRVDIDMDQPKSTF---KKADGSEVSEFLEYRKQYNOEITDLKQ 273
 264 KVEVTHGQGMGRKRVCAVTRPASHQTFPLQESQGVETVAQYKQKYNQ--LKY 320
 274 PVL-----VSQPRRRRGGTLPGPAMLIPELCYLTG-----LTKKRDENVMKDLAV 322
 321 PHLPCLAGQGEQKH-----TYLPLEVCMNVAQGRICIKKLTID--NQTSTM--IKA 365
 323 HTLRLPEGRORVGLIDYIHKNQNVORELDMGSLFSDNLSGSGILQTEKHOGGKT 382
 366 TARSAPD--KQEBISLHK--NASNYLDPIYIEFGIKVMDMTVETGVLPAPIIQYGRN 422
 383 PDY--NPQPADWSKETRGAPLISVXPLDMLIYFRNYSANSLIQ---LFTVTPAMG 437
 423 RAIAIPENQGVW--DMRGKQFYNGEIKYMAIACFAPQCKREEVILKNFTDLRKISDAG 480
 438 MQMRK-----ALMIEVDDBTEAVRLVLOQKVTADQIVVCLLSNRKQYDAIKKYLCTDC 493
 481 MPIQQCFCKYKQADQADVEPMFRHLKNTYSG--LQILIVIL--PKGTPYIAVEKVGDTLL 538
 494 PTPSQCVARTLGKQQTMAIATKIALQMCNCKMGELWRVDP-----LKVMTVIGID 546
 539 GNATQCVQKVVVK--TSPQTLNCLKINVLGG--INNIIIVPQRSVAFQOPIVIFGLAD 595
 547 CYHMDTGA---RSLAGFVVASINEGMTWFSRCIFQDQSGVLDGKVCQALALRANSC 603
 586 VTHP--PADGKKRPSITAVAGSMDAHPSRYCATVVRQPROSITIELSLTWAEILLIORYKS 654
 604 NEYMSRRIIVYRVDGVDGQDLKTLVNYEVEPOFLDCLKSIGRGVNPRLTVIVYKKEVNRFF 663
 655 TRFKETRIIFVYRVDGVBEGQLPQILHYELLARDACTIKLEKDYQPGITIVVQKHHRTLF 714
 664 A---QSGRLONLPFGIVIVETVREPMYDFPIYSQAVRSGSVSPFTYNTYINSGLKP 719
 715 CADKNERIGKSGNIPAGITVDNITHPREFBYILCSHAGIQTSRSHYVLMDDNRTFA 774
 720 DHIORLTYLCHIIYVWMPGVIVRPAFCQYANHLAF-----LVGOSIH 761
 775 DELQILITVGLCHTYVRCRVSISIPAPAYARLVAVRARYHLVKEHDSGESHISGQNG 834
 762 REP 764
 835 RDP 837

RESULT 6
 12C2 RAT ID 12C2 RAT STANDARD; PRT; 860 AA.
 AC 0902EL;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
 GN eIF2C2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
 RX MEDLINE=99443791; PubMed=10512872;
 RA Chakaluk D.E., Tabhaz N., Hendricks L.C., Dimattia G.E., Hansen D.,
 RA Pilgrim D., Hodman T.C.;
 RT "GERP95, a membrane-associated protein that belongs to a family of
 RT proteins involved in stem cell differentiation.";
 RL Mol. Biol. Cell 10:3357-3372(1999).
 CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
 CC initiation process.
 CC -1- SIMILARITY: belongs to the argonaute family.
 CC -1- SIMILARITY: contains 1 PAZ domain.
 CC -1- SIMILARITY: contains 1 P1W1 domain.
 CC -1- SIMILARITY: contains 1 P1W1 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AF195534; AAF12800.1; ALT_INIT.
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; P1W1.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1W1; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; P1W1; 1.
 KM Initiation factor; Protein biosynthesis.
 FT DOMAIN 226 349
 FT 518 819 P1W1.
 SQ SEQUENCE 860 AA; 97318 MW; A5B0798C66481C9C CRC64;

Query Match 12.3%; Score 504.5; DB 1; Length 860;
 Best Local Similarity 24.1%; Pred. No. 1.7e-28;
 Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;

20 GSSGIVRLSTNHFRLTSRPMALYOHYHIDYPLMEARRLSAL--LFQH--EDLIG-K 73
 33 GTTGRTIKQANFFEM--DIPKIDYHYELDIPKDCPRVNVNIEYHVVQHFKQIFGDR 91
 74 CHAFDQ-----TIFLPKLOQKTVESKTRNGED--VRIT-----T 110
 92 KPVFDRKRLVYAMPPIG--RDKVELLEVTLPSGKRIKIVKISCVSLQALHDALSG 150
 151 RLPSVFEPTIQALDVAMRH--LPSMKRTYVGRSFTLASBSCNPLDGGGRVWGFHQSVP 209
 114 ELPEPTSPCLQFYNIIFRLIKIMLQIGHYVNPNDPIDIP--SHRLVIMPGETTSILQ 172
 173 YENSIMLCTDVSHKVL--RSEVLDPMF-----NFTQTEHKEK--FQGVSKELIGLVYL 223
 210 SLWKMLNIDVSATAFYKQPIVIEFVCEVLDIFSIEQKPLTDSQKRVFTKIKGLKVB 269
 224 TKY--NKKTYRVDIDMDQPKSTFKADGS---EVFLEYRKQYNOEITDLKQPYL 276
 270 ITHGGMKRTKRVCAVTRPASHQTFPLQESQGVETVAQYKQKYNQ--LKY 329

QY 277 -VSQPKRRRGPGGTLPGPAMLIPELCYLTG-----LTDKMRDENVMDLAVHRLTP 328
DB 330 QVGEQCKH-----TYLPLEVCNIVAGQRCIKKLTLD---NOTSTM--IRATASAP 374
QY 329 EQRGREVRGLDIYHKNDNVQRELRDWGLSDNLSPSGRILOTEKHOGKTFDVPQ 388
DB 375 D-ROEELSKLMR--SASFNTPYRBEQIMVKDMTDTGRLVQPSILYGR-----424
QY 389 FADMSKETRGAPLISVKELDN-----MLLITRREYEA---NSLIQNLKVT 433
DB 425 -----NKAIATPVQGVMDMRKQFHTGIEIKVMAIACAPQRCQTEVHLKSFTEQLKIS 479
QY 434 PAMGQMRK-----AIMIEVDRTYALRVLOQKXTADTQIVCLSSNRKDXDAIKKYL 489
DB 480 RDAGMPIQGQPCFKYAGADSVEMFPHL-KATYAGIQLVVVL-PGKTPVYAEVRVG 537
QY 490 CTDCPTSPQCVARTLGKQOTVMAIATKIALQNNCKMGG-----ELMRVDIPLKLVMIY 543
DB 538 DTVMGATQCVQGMKV--QRTTPQTLNLCIKINVKLGNNILLPGQRPVFGQVIFL 595
QY 544 GIDCHDMTAG--RRSIAGVYASINSGMTWFSRCIFQDRGQELVDGLKVCLOALRAW 600
DB 596 GADVTHP-PADGKKPSIAAVGSDAFNRYCATVRVQGRHBIODLAMVRELLIOF 654
QY 601 NSCNEVMSRIIVYRDVGDDGQDKTLVYEVPOFLDKLSIGRGNPRLTVIVKKEVNT 660
DB 655 YKSTRTFKETRLIFRDVSEGOVQLHLLAIREACIKLEKYOIGITITIVQKXHT 714
QY 661 RFFA-----QSGRLQNPFGTYIDVETRPETVFFIVSQAVRSGSVPTHYNTVNSG 716
DB 715 RLFCFDKNERVKGKSNIPAGTIVDTKITHPTPEFDPYLCSHAGIOGTSRPSHYHLMWDMR 774
QY 717 LKPDHIGRLTYKLCIYYNMFGVIRVAPQYAKHLAF 754
DB 775 FSSDELQILTYQLCHTYVRCRSVSIPAPAYYALHVAR 812

RESULT 7
ID 12C4_MOUSE STANDARD; PRT; 861 AA.
AC Q8CQF8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 4 (eIF2C 4) (eIF-2C 4)
DE (P1w1/argonaute family protein meIF2C4).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Doi N., Zemo S., Ut-Tel K., Takahasi F., Ueda R., Miyata Y.,
RT Saigo K.;
RT "eIF2C family proteins and Dicer homologs are required for siRNA
mediated RNAi in mammalian cells";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process (By similarity).
CC - SIMILARITY: Belongs to the argonaute family.
CC - SIMILARITY: Contains 1 PAZ domain.
CC - SIMILARITY: Contains 1 P1w1 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).

DR EMBL AB081474; BAC15769.1; --
DR MGI:1924100; E1F2C4.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; P1w1.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; P1w1; 1.
DR PROSITE; PS00821; PAZ; 1.
DR PROSITE; PS00822; P1w1; 1.
KW Initiation factor; protein biosynthesis.
FT DOMAIN 225 338
FT PAZ.
FT DOMAIN 509 820
FT P1w1.
SQ SEQUENCE 861 AA; 97006 MW; 791F510B220E389 CRC64;
Query Match 12.3%; Score 504.5; DB 1; Length 861;
Best Local Similarity 24.2%; Pred. No. 1.7e-28;
Matches 207; Conservative 146; Mismatches 358; Indels 143; Gaps 34;
QY 20 GSGGIIVRLSTNHFRLTSRPPNALYQYHIDVPLMEARLRSLAL-----FQHEDLICK 73
DB 22 GTVGKEIRLLAHFQV-QIPKIDVYHYVDIDPEKRPFRVNEVVDVTRVHFRKQIFGDR 80
QY 74 CHAFDG-----TLFPLP---KRLQOKTEVFSKTRNGEDVRIITLV-----111
DB 81 QPQYDGRKRMVTHAPLPIGRDIDMEVT---LPBGKQOTKRVSVQWVSVALQLLEA 136
QY 112 ---TNELEPTSPYCLQFYNIIFRLKIKMLQOIGRNTYNNPNDPIDP-SHRLVYWPGE 166
DB 137 LAGHLNEVPDDSVQALD---VITRHLPEMRVTPVGRSPFSPDEGYHPLGGGREVWFGF 192
QY 167 TTSILQYENSIMLCIDVSHKL-RSETVLDME-----NFMQTEBHKFOEQV--SKEL 217
DB 193 HOSVRPAMNMMLNDIVSATAFYRAOPTIEFMCEVLDIQNEQKPLDSQVKEVTKGI 252
QY 218 IGLVVLTKY---NNKTYRVDDIDWDQNPSTF---KKADGSEVSFLERYKQNOEITD 270
DB 253 RGLKVEVTHCGQKRRKRYRCNVTTRPASHQTFPLQENQAMECTVAQVFKQKYSQLYKH 312
QY 271 LKQPVV-VSQPKRRRGPGGTLPGPAMLIPELCYLTG-----LTDKMRDENVMDLAV 322
DB 313 PHLPLQVQGEQGR-----TYLPLEVCNIVAGQRCIKKLTLD---NOTSTM--IRA 357
QY 323 HTRLTEPQORQVRYGLIDYHKNDNVQRELRDWGLSDNLSPSGRILOTEKHOGKKT 382
DB 358 TIASAPD-ROEELSRIVKNSNWGSDPDLKEGIVVHNMELTGRVLPAPLQVGRNV 416
QY 383 FDY-NQPADWSKETRGAPLISVKELDNMLITRYNRYEASNLION-----LEKTPANG 437
DB 417 KTVATPSQGVW--DMRGKQFYAGIEIKVWAVACAFQKOCREDLKSFTDQLEKISKDA 474
QY 438 MOKRK-----AIMIEVDRTYALRVLOQKXT-ADTQIVCLSSNRKDXDAIKKYLCTD 492
DB 475 MPIOGQPCFKYAGADSVEMFPHL-KATYAGIQLVVVL-PGKTPVYAEVRAGDIL 531
QY 493 CPTSPQCVARTLGKQOTVMAIATKIALQNNCKMGBELWEVDIP-----LKLWITVG 545
DB 532 LGMATQCVQIKNVK--TSPQTLNLCIKINVKLGNNILLPGQRPVFGQVIFLGA 588
QY 546 DCHDMTAG--RRSIAGVYASINSGMTWFSRCIFQDR-----GQELVDGLKVC 592
DB 589 DVTHP-PAGGKKPSIAAVGSDHPSKICATVWQTSQELAQOELLYSQEVVQDLTSM 647
QY 593 LQALRAMNSCNEVMSRIIVYRDVGDDGQDKTLVYEVPOFLDKLSIGRGNPRLTVI 652
DB 648 ABELLIFYKSTPFKPRITIIYVAGVSEGMQVAPBELAIRKACISLEEDRPGLTYI 707
QY 653 VYKRVNTRPFA-----QSGRLQNPFGTYIDVETRPETVFFIVSQAVRSGSVSPTHY 708
DB 708 VVQGRHHTRLFCADKMERVKSGVNPAGTIVDTSTVHPSFDYLCSHAIGQTSRSHY 767
QY 709 NVIYDNGSLKPDHIGRLTYKLCIYYNMFGVIRVAPQYAKHLAF-----754
DB 768 QVLMDDNCFLADEQLLTYQLCHTYVRCRSVSIPAPAYYALHVARVHLYDKDHSA 827

QY 755 ----LVGQSHREP 764
DB 828 EGSVHSGQSGNRDP 841

RESULT 8
ID 12C2_RABIT STANDARD; PRT; 840 AA.
12C2_RABIT
16-OCT-2001 (Rel. 40, Created)
15-MAR-2004 (Rel. 43, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE (Fragment).
GN EIF2C2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=98267198; PubMed=9602122;
RA Zou C., Zhang Z., Wu S., Osterman J.C.;
RT "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL Gene 211:187-194(1998).
CC - FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process.
CC - SIMILARITY: Belongs to the argonaute family.
CC - SIMILARITY: Contains 1 PAZ domain.
CC - SIMILARITY: Contains 1 PAZ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by, and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).CC - EMBL AF005355; AAC24323.1; ALT_INIT.
DR PIR; PC0505; J05659.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003105; PAZ.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PAZ; 1.
DR PROSITE; PS00821; PAZ; 1.
DR PROSITE; PS00822; PAZ; 1.
KW Initiation factor; Protein biosynthesis.
FT NON_TER 1
FT DOMAIN 216 329 PAZ.
FT 498 799 PIWI.
SQ SEQUENCE 840 AA; 95305 MW; 1E703F9E31391F29 CRC64;

Query Match 12.3%; Score 503.5; DB 1; Length 840;
Best Local Similarity 24.1%; Pred. No. 2e-26;
Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;

QY 20 GSSGIVRLSTNHFRLTSRPPALVQYHIDNPLMEARLRSL--LFQH--EDLIQ-K 73
DB 13 GSGRTIKLQANFFEM-DIPKIDIDHYEELDIKBEKPRVNEIYEHVQHFKAQIFDR 71
QY 74 CHAFDQ-----TLFPRKLOOKVTEVFSTRNGED--VRIT-----TLTN 113
DB 72 KPFVDFGRKULYTMPPIG-REKVELLEVLPBEGKRIKVKVIXKWSVCVSLQALHDAISG 130
QY 114 ELFPTPTLQFYNIIFRLIKLWNLQIGRNYNNDPIDIP-SRLVYNGFTTSIIQ 172
DB 131 RLTSVVFETIQLADVVRH-LPSMRATPVGRSFFTSBSCSNPLGGREWGFQSVRP 189
QY 173 YKSNIMCTDVSHKVL-RSETVDF--MNFYHQTSEHK-----FOEVSKEILGLVYL 223
DB 190 SLWKMLNIDVSAATFYKQPVLEFVCEVLDFKSIEQKPLTDSQKRVKFTTEIKGLKVE 249

QY 224 TKY---NKTAYVDIDMDQNKSTPKKADG-----EVSFLEYRKYNNQELTDLKQPV 276
DB 250 ITHGQMKRKYRVNCTVRRPASHQTFPLQESGQTEVCTVAQYFDRHCLVIRPHLPCL 309
QY 277 -VSQPKRRRGGTLLEGFAMLIPELCYLTG-----LTKRKNDFNNKDLAVHTRLP 328
DB 310 QVGEQKH-----TYLPLEVCNIVAGRCIKLID--NGISTW--IRATARSAP 354
QY 329 EQRQREVGKLIYHKNVQRELPDWSLSPSNLSFSRILLOEKIHQSGKTFDYNQ 388
DB 355 D-ROEISIKLMR--SASFITDYPVHFGIMVDEMTDVTGVLQPSIYGR----- 404
QY 389 FADWSKTRGAPLISVPLDN-----MLIYTRNYEAA--NSLIQNLKVT 433
DB 405 -----NKATATPVQGVMDNRNKGPHGIEIKYMALACFAPQRCQTEVHLKSFELRKIS 459
QY 434 PANGWQMRK--ALMIEVDRTAYLVLOQKVTADTVQVCLSSNKKDKYDAIKKL 489
DB 460 RDAGMPIDQGPCKYAGADSVGPMFRHL-KNTYAGQLVAVIL-PGKTPYAEYKRVG 517
QY 490 CTDCPTSPQCVARTGKQOVTWALATKIALQNCMG-----ELMREVDIPKLWTV 543
DB 518 DTVLGMAITQCVQKMN--QRTTPQILSNCLKINVLGGVNNILLPQGRPFVFGQVIEL 575
QY 544 GIDCVYDMTAG--RRSIAGFVASINEGTRMFSRCIFDRQGLVDGLKVCLOALRLW 600
DB 576 GADVTHP-PAGDCKKSIIVAVGSMDAHNRKCATVRYQVQHQEIIQDLAMVRELIQF 634
QY 601 NSCNEVMPRIIYRPGVGDGQKTLVYEVDFDCLKSIGSGYRPLTVYKRVNT 660
DB 635 YKSTRKPRILYRGGVSEGOQVYHHELAIRACIKLEMDYDQGLTFIVQGRHNT 694
QY 661 RPPA---OSGRLQNPRLPGTVIVETVREYDFVTSQAVRSGSVSTHNVITDNG 716
DB 695 RLCTDKNERVSGNIPAGTVYDTKITRPTEDDFYLSHAGIQGTSRSHVHWDNR 754
QY 717 LKPDHQRILYKLCIHYMMPGVIRVAPCOVAKLAF 754
DB 755 FSSDEQLILTYQLCHYVTRCTSVSIPAPAYIAHLVAF 792

RESULT 9
PINH_ARATH STANDARD; PRT; 988 AA.
ID PINH_ARATH
AC Q9XGM1; O49256;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PINHEAD protein (ZWILLR protein).
GN AT5G43810 OR MOD19.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=9094873; PubMed=9876176;
RA Lynn K., Fernandez A., Alda M., Sedbrook J., Tasaka M., Mason P.,
BA Barton M.K.;
RT "The PINHEAD/ZWILLR gene acts pleiotropically in Arabidopsis
development and has overlapping functions with the ARGONAUTE1 gene.";
RL Development 126:469-481(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=96169396; PubMed=9501101;
RA Mousavian B., Schoof H., Haeccker A., Uerogens G., Laux T.;
RT "Role of the ZWILLR gene in the regulation of central shoot meristem
cell fate during Arabidopsis embryogenesis.";
RL EMBO J. 17:1799-1809(1998).

[3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RA Kaneo T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
 RL Submitted (Apr 1999) to the EMBL/GenBank/DBS databases.
 CC -1- FUNCTION: Developmental protein required for reliable formation of
 CC primary and axillary shoot apical meristems. May be a component of
 CC a hypothetical meristem forming competence factor.
 CC -1- SIMILARITY: Belongs to the argonaute family.
 CC -1- SIMILARITY: Contains 1 PAZ domain.
 CC -1- SIMILARITY: Contains 1 P1w1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF154272; AAD40098.1; -
 CC EMBL; AJ223508; CA11429.1; -
 CC EMBL; AB026551; BAB11310.1; -
 CC PIR; T52134; T52134.
 CC DR PIR; T52134; T52134.
 CC DR InterPro; IPR003100; PAZ.
 CC DR InterPro; IPR003165; P1w1.
 CC DR Pfam; PF02170; PAZ; 1.
 CC DR Pfam; PF02171; P1w1; 1.
 CC DR PROSITE; PS50821; PAZ; 1.
 CC DR PROSITE; PS50822; P1w1; 1.
 CC DR Developmental protein.
 CC KW DOMAIN 337 451
 CC FT DOMAIN 625 946
 CC FT CONFLICT 475 479 PRDRE -> AEGQR (IN REF. 2).
 CC FT CONFLICT 671 671 N -> D (IN REF. 2).
 CC SQ SEQUENCE 988 AA; 110867 MW; 32EBB349C61DA20 CRC64;
 CC -----
 CC Query Match 12.3%; Score 503; DB 1; Length 988;
 CC Best Local Similarity 23.7%; Pred. No. 2.7e-28;
 CC Matches 204; Conservative 136; Mismatches 399; Indels 122; Gaps 32;
 CC -----
 CC 5 VNTRONLHVHESKSGSGLIVRLSTNFRSLRSPQWALYOHIDYDLMARSLRL- 63
 CC 120 MGVNRKNSNAPRPGGTIGTKIVKANHF-LADFTKOLNDYDVTITEVSSKSNRLII 178
 CC 64 -----LFOEDLIGKCHAFD-----GTILFLPKLQCKYTE---VPSKTNGEDVR 106
 CC 179 AELVRLYKESDGRRLPAYDGRKSLYTAGELPFTWKEFSVKIYEDDDIINGPKERSYK 238
 CC 107 IIT-----LTNELPTSPICLOFYNIIFRLKIKIMLQOIGNYVNPNDPI 153
 CC 233 VALKVARANMHLSEFLAGKADCPQBAVQILDIVLRE-LSVKRCFVGSGSPSPD-I 295
 CC 154 DIP---SHRLVTMPFTTSIIQYENSIMLCYDV-SHKVLRSETVLDFMNFYHOTEHK- 208
 CC 296 KIPQRLSEGLSEWCGFYQSIPTQGLSINIDMAAFIEPLPVEFAQLIGKDVLSKP 355
 CC 209 -----FOEVSKEELGLVLTKN---NKTFRYDQIDWQONPKSTFKADGSEV-SLEFY 260
 CC 356 LSDSRVAKIKGLRGVKEVTHRANVRKRIRVAGLFTQPTRELMFPVDENCTKASVIRY 415
 CC 261 RKOYNOBITDLKQPYL-VSQPRRRRPGGLTLPGLMILPELCYLTGLTDKXRNDFNWKD 319
 CC 416 GMYGFTIHTLPCLOVGNCK---ASYLPMWACKIIVEGQRYT---KELNEKQITVL 467
 CC 320 LAVHRLTPEQOREVGRILDIYHNKDNVQRLRWGSLFSDNLSFGSLIQTE--KIH 377
 CC 468 LKV-TCQRRDRDENILRTVQ--HNAYQDDPYAKSEFGNISEKLASVEARILPAPWLKTH 524
 CC 378 OGKTFDYVPOPADSKETRGAFLISVKPLDWMLIYTRANE--AANSLQNTLFTKTPA 435

Db 525 ENGEKDCDLPQVQGMNMMNK--IKINGMTVSRNACVNSFSVQENARFGCNELQGWCEV 562
 Qy 436 MGNOMRAIMIEV-----DPRTE-----AYDRLQOKVTADTQIVCLLSNRKDYDAIK 486
 Db 583 SGMEFNEPEVPIPIYSRPPDVEKALKHVHTSNKTKGKELELLAILPNNNSISVGDLE 642
 Qy 487 KYLCTDPTSPQCVAR--TLKQOTWALATKIALQNNCKNGE-----LNRV-- 533
 Db 643 RICETELGLISQCLTGHVRIKQO-----YLANVSLKINVKNGRNTVLVDAISCRIP 697
 Qy 534 --DIPPLVMIWIDVCHMTA--GRRSIAGFVASIN--EGMTWPFRCIFQDQGOELVDG 588
 Db 698 VSDIP--TIIFADVTHPENGESESSPSIAVVASQDWEVITKXAGLVCAQARQELIQD 754
 Qy 589 LKVCLOALR-----AMNSCNEVMSRIIVRYDGVGDQQLKLVNYEVPQFL 635
 Db 755 LYKWDOPVAGTVSGGMRDLISFRKATGQKRLIIFRYDGVSEGFVLLYEYDAIR 814
 Qy 636 DCLKSIKRGYNPLTYIVYKKRVYTRFFA-----QSGGLQPLGTVIVDVEVTRPEWY 689
 Db 815 KACASLEPNYQPVPTPIVQKRRHTRLPANNHRDKNSTDSGNILPGTVVDYIKIHPTEF 874
 Qy 690 DFFIVSQAVRSGSVPTNYVIVYDNGELKPDHIOQLTYKLCIHYNMPGVIRVAPCOYA 749
 Db 875 DFYLSHAGIGQTSRPAHYVLMDENNFTADGIQSLTNLCYTYARCTRSVSIVPAYYA 934
 Qy 750 HCLAFVVGSIHREPNLSLN 770
 Db 935 HLAFAFR--RFYLEPEIMQDN 953

 RESULT 10
 ID 12C3_MOUSE STANDARD; PRT; 860 AA.
 AC Q8CJF9;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Eukaryotic translation initiation factor 2C 3 (eIF2C 3) (eIF-2C 3)
 DE (P1w1/argonaute family protein meif2C3).
 GN EIF2C3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_Taxid=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Doi N., Zemo S., Yi-Tei K., Takahashi F., Ueda R., Miyata Y.,
 RA Saigo K.;
 RT "eIF2C family proteins and Dicer homologs are required for siRNA
 RT mediated RNAi in mammalian cells."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBS databases.
 CC -1- FUNCTION: plays an important role in the eukaryotic peptide chain
 CC initiation process (by similarity).
 CC -1- SIMILARITY: Belongs to the argonaute family.
 CC -1- SIMILARITY: Contains 1 PAZ domain.
 CC -1- SIMILARITY: Contains 1 P1w1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB081473; BAC15768.1; -
 CC MGD; MGI:2140548; AM048686.
 CC DR InterPro; IPR003100; PAZ.
 CC DR InterPro; IPR003165; P1w1.
 CC DR Pfam; PF02170; PAZ; 1.
 CC DR Pfam; PF02171; P1w1; 1.
 CC DR PROSITE; PS50821; PAZ; 1.

Wed May 12 09:50:22 2004

us-10-043-774b-2.rsp

Page 10

DR PROSITE; PS50822; PIWI, 1.
KM Initiation factor; Protein biosynthesis.
FT DOMAIN 236 349 PAZ.
FT DOMAIN 518 819 PIWI.
SQ SEQUENCE 860 AA; 97240 MW; 31215536BF84BA61 CRC64;

Query Match 12.2%; Score 502; DB 1; Length 860;
Best Local Similarity 24.4%; Pred. No. 2.6e-28;
Matches 209; Conservative 147; Mismatches 349; Indels 152; Gaps 37;

QY 20 GSSGIVLSTNHPILTRPQWALYQYIDYNPIMLMEARRLSALLPQHEDIGCHAFDG 79
DB 24 GIMGKPIKLANGFOV-EIPKIDVLYEVDIKPKCPRRV-----NREVVDSKVQHFV 76
QY 80 TI-----LPLKRLQOKVT-----EVSSTRNGED-----VRITITLT 112
DB 77 TIFGDRFVYDOKRSLYANPLPVATGVDVTLPGGKDRPFKVSVKFSRVSWHL 136
QY 113 NE-----LP-----PTSPCLQFNIIIFRLLKIMNLQIGRNYNPNDDIP-SH 158
DB 137 HEALAGTLPBPLDCKVSTNPPHADVLRH-LPSMKYTPVGRSFFSAPEGYDPLGG 195
QY 159 RLVIWPGFTTSILQYENSIMLCTDVSHTVL-RSETVLDFM-----FNFYHQ-----TEEH 207
DB 196 GREVWFGHGSVRPAMWGMNIDVSATAFYKAPVIOFMCVLDINIDEGPRPLTDSH 255
QY 208 KFGQGVSKELIGLVVLTXY---NNKTYRVDIDWDONPKSTF---KKADGSEVSLEY 260
DB 256 RV--KFTKEIKGLKVEVTHGCTMRKVCNVTERRPASHQTFPQLNRGQVETVAQVF 313
QY 261 RKOYQCEITDKQOPVL---VSOPKRRRPGGTLPGRMILPELCYLTG-----LTDK 309
DB 314 REKYLQ---LKYPIPLCLOVQGEKX-----TYLPLECNIVAGRCIKLTD- 359
QY 310 MENDFNWKDLAVHTRLTPEQORQREVGRLIDYHNKNVQRE-LRDMGLSFDNSLISFS 367
DB 360 --NOTSTM--IKATARSAPD-ROEISRLV---RSANVEYEDPVOEFOLEVRDEMAHVT 410
QY 368 GRILQTEKHOGSKTFED--NPOFADWSKETEAPLISKPLDNMLL--IYTRR--NTEAA 422
DB 411 GAVLPAPMLQYGRKRTATATPSGYW--DNEGKQHEGVEYKMAIACFATQRCREIL 468
QY 423 NSLIQNLKRYTPAMGMQNRK---AIMEVDRTREALRVLQOKVTADTOIVCLLSNR 478
DB 469 KGFDTQLRKISKDAGMPIQGPFCFKYAGADSVEMPRHLKNYSG-LQLIIVIL--PGK 526
QY 479 KDKYAIKYLCTDPTSPSCVNAVATLGKQOTWALAKIALQNNCKMG-----ELMR 532
DB 537 TPVNAEVRVADPTLLGMATQCVQVNVK--TSPQTSNLCLKINVLKGGNNILVPHQR 584
QY 533 VDIPLKVMIVGIDCYHDMTAG---RRSIAGFVASINEGTRWFSRCIFQDRGOELVGL 589
DB 585 PSVFQQPIVIFGADVTHP--PAGDGKPSIAAVGSMADHSPRYCATYRVGQPRDEITODL 643
QY 590 KYCLQALRAMNSCNEYPSRIIVYRDVGDQKLTLYNVEVPQDLCKSIGGYNPL 649
DB 644 ASMWELLIOFKSTRFKPRTIIFRSDVSSEGQFOVLYELLAIRESCLISLEQYRPGI 703
QY 650 TVIVVKKKVNTPFF---AOSGRLQNPPLPGTVIDEVTREPMYDFIVSOAVNSGVS 705
DB 704 TVIVVQKHHHTLPCADRTERRVGRSGNIPAGITVDTDTTHYERDFYCSNAGIQGNRP 763
QY 706 THYANVYDNGSLKPHIQILTYKLCIITYNMPGVIRVAPCCYAKLAF-----754
DB 764 SHYHVLMDNFTADDELQILTYQLCHTYVRCRVSIPAPAYVYALVAFRAPHYLVKHE 823
QY 755 -----LVGQSIHREP 764
DB 824 DSAEGSHVSGQSGNDP 840

RESULT 11
12CL_MOUSE

ID 12CL_MOUSE STANDARD; ERT; 857 AA.
AC O8CUG1;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 1 (EIF2C 1) (EIF-2C 1)
GN EIF2C1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RF Doi N., Zemo S., U-Tel K., Takahasi F., Ueda R., Miyata Y.,
RA Saigo K.;
RT "EIF2C family proteins and Dicer homologs are required for siRNA
mediated RNAi in mammalian cells";
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process.
CC - SIMILARITY: Belongs to the argonaute family.
CC - SIMILARITY: Contains 1 PAZ domain.
CC - SIMILARITY: Contains 1 PIWI domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
or send an email to license@isb-sb.ch).
CC
DR EMBL; AB081471; BAC15766.1; -.
DR MGD; MG1; 2446630; EIF2C1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
FM Initiation factor; Protein biosynthesis.
FT DOMAIN 226 346 PAZ.
FT DOMAIN 515 816 PIWI.
SQ SEQUENCE 857 AA; 97101 MW; 940838503153A5E2 CRC64;

Query Match 12.2%; Score 499.5; DB 1; Length 857;
Best Local Similarity 24.0%; Pred. No. 3.9e-28;
Matches 202; Conservative 160; Mismatches 348; Indels 133; Gaps 37;

QY 20 GSSGIVLSTNHPILTRPQWALYQYIDYNPIMLMEARRLSALLPQHEDIGCHAFDG 79
DB 30 GTVGRKIKILANVFEV-DIPKIDVHYEVDIPDCPRVNEVEYVWQHFKPOIFGDR 88
QY 74 CHAFDG-----TILFPL---KRLQKVEVFSKTINGED---VRITITLT-----T 110
DB 89 KPYVDKENVITYVTALPIENKRVDEVT---IPBGRDRITKVISIKMLATYSWMTHRA 144
QY 111 LTNELPPTSPCLQFNIIIFRLLKIMNLQIGRNYNPNDDIP-SHRLVIWPGFTTS 169
DB 145 LVSGQIFVPLESQAQDVMRLIAS-MRYTPVGRSFFSPPEGYVHPLGGREWFGFHQS 203
QY 170 ILQYENSIMLCTDVSHTVL-RSETVLDFM-----FNHYQBEKKFOGV--SKELIDL 220
DB 204 VRPAMWGMNIDVSATAFYKAPVIOFMCVLDINIDEGPRPLTDSGRVAFTEIKEL 263
QY 221 VVLTKY---NNKTYRVDIDWDONPKSTF---KKADGSEVSLEYRQYNOEITDKQ 273
DB 264 KVEVTHGQMKRKYRVCNVTERRPASHQTFPQLNREGQVETVAQHFKQKYLQ---LKY 320
QY 274 PVL---VSOPKRRRPGGTLPGRMILPELCYLTG-----LTDKMDNFNWKDLAV 322
DB 321 PHLPCLQVQGEKX-----TYLPLECNIVAGRCIKLTD---NOTSTM--IKA 365


```

QY 323 HTRLRPEQREVRGLIDYTHKNDVQRLRDMGLSPDSNLSFSGRLIQTEKHQGKT 382
DB 366 TRKASAD-RQBEISRLMK--NASCNDPIYQFEGIKVODMEVGRVLPAPILQYGRN 422
QY 383 PDY-NPOFADMSKETRGAPLISVPLDMLLYTRRNEAANSLIQN--LFKVTAMG 437
DB 423 RAIAFENQGVW--DMRGKFYNGIEIKVWAIACFAPKQCRREELKFNFDLRKISKDAG 480
QY 438 MGRK-----AIMIEVDRTREALYLQCKVADPQIVVCLSSNKKDYDAIKKYLCTDC 493
DB 481 MPIOGPFCCKYKAGADSVEMFRLKNTYSG-LQLIIVIL--PGKTPYAEKVRGDTLL 538
QY 494 PPSQCVVARTLGKQQTWAIATKIALQNMCKMGELEMRVDIP-----LKLVMIVGID 546
DB 539 GMATQCVYKNAVK--TSPQTLISNCLKINVLGG--INNILVPHGRSAVFOQPVIFLGAD 595
QY 547 CYHDMTAG--RRSIAGFYASINEGNTWFSRCIFQDRGQELVDGLKVCLOALPAMNSC 603
DB 596 VTHP-PAGDGKPSITAAVGSMDAPSRXCATVRQRPQOEIIEDLSYWEELLIOFYKS 654
QY 604 NEMPSRIIVYDGVGDQGLKTLVNYEVPQFLDCKSIGRGVNPRLTVIVKRYNTRFF 663
DB 655 TRFKPRLIFRYDGPBEGQLPOLHIELAIDACIKLEKDYQPIITIVQKRHTLFL 714
QY 664 A-----QSGRLQNPFGVIVDVEVTRPEWDFPFIQSAVRSVSPTHYNYLYNSGLKP 719
DB 715 CADKRIKSGNIPAGITVDNTNTHPEFDFYLCSHAGIQGTSRPHSYVYLMQDNRETA 774
QY 720 DHIOGLTYLCTGIYNNWBEIVVAPCOYAKHLAR-----LVGSGIH 761
DB 775 DELQILTYOLCHTYRCTRVSIPAPAYARLVAFRARIYLVKDHSGEGSHISGSGNG 834
QY 762 REP 764
DB 835 RDP 837

RESULT 12
1203 HUMAN STANDARD; PRT; 860 AA.
AC 09H9G7; 09H1U6;
DT 15-MAR-2004 (rel. 43, Created)
DT 15-MAR-2004 (rel. 43, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 3 (eIF2C 3) (eIF-2C 3)
DE (Argonaute 3).
GN EIF2C3 OR AGO3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN 11]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto T.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masudo Y., Oshima A.;
RT "NEO human cDNA sequencing project."
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN 12]
RP SEQUENCE OF 221-860 FROM N.A.
RA Brown A.;
RT Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (by similarity).
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AK022827; BAB14262.1;
DR EMBL; AL138787; CAC21457.1;
DR GenBank; HGNC:18421; EIF2C3.
DR MIM; 607355;
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; Piwi; 1.
KM Initiation factor; Protein biosynthesis.
FT DOMAIN 236 349
FT DOMAIN 518 819
FT SEQUENCE 860 AA; 97330 MW; BED26B529A0CF920 CRC64;
SQ
Query Match 12.1%; Score 497; DB 1; Length 860;
Best Local Similarity 24.3%; Pred. No. 6e-28;
Matches 207; Conservative 150; Mismatches 354; Indels 140; Gaps 37;
QY 20 GSSGIIVRLSTNHRFLTRSPQWALYOHIDYNPLMEARRLSAL--LFQHEDLI--GK 73
DB 24 GAMGKPIKLANFOV-EIKIIVLYEVDIKDKCPRRNVREVDMSVGHFVYITFGDR 82
QY 74 CHAFDG-----TLFLPKRLQCKVTEFSTRNGED-----VRITITLTNE----- 114
DB 83 RPYVDGRSLXYTANPLVATVATGVDLDTLPBEGQKDRPFVSKYFSRVSWMLHLEVLG 142
QY 115 --LP-----PTSPICLQFYNIIFRLIKINMLQOIGNYYNPNDIDIP-SHRLVIWP 164
DB 143 RTLPPELELDKPISTNVHAVDVLRH-LPSMKYTPVGRGFFAPBEGYDHPGLGGREWF 201
QY 165 GFTTSIQYENSIMLCVDVSHKYL-RSEYVLDMY-----FNFYHQ-----TEHKKPQEOV 213
DB 202 GFHQSVPRAMKMKMLINDVSATAPFYKAQPIVQICMEVDLIDHIDEQRPILDSHRV--KF 259
QY 214 SKELIGLVLTFRY--NNKTYRVDIDPDQNPSTF---KKADGSEVSELEYRKOYQ 266
DB 260 TKEIKGLKAVTHGCMRKRCYCNVTRPARSHQTPFLQENQGVETRVAAQFRKCYTL 319
QY 267 EITDLPQVLT---VSQPKRRRPGGTLPGPAMLIPELCYLTG-----LTDKMRDYN 315
DB 320 Q--LKYFHLPCLOVQGEQGH-----TYLPLEVCNIVAGRCIKLTD--NOTS 363
QY 316 VMKDLAVHTLTPGEORREYGRLLIDYTHKNDVQRL--LDMGLSPDSNLSFSGRLIQ 373
DB 364 TW--IKATASAD-RQBEISRLV--RSANETDPFQEPFKRDEMAHATGRVLP 416
QY 374 EKHQGGKTFDY-NPOFADMSKETRGAPLISVPLDMLLYTRR--NYEANSILQ 428
DB 417 PMLQYGRNNTVATPSHGVW--DMRGKFYNGIEIKWAIACFAPKQCRREELKFNFDL 474
QY 429 LFKVTAMGMQMR-----AIMIEVDRTREALYLQCKVADPQIVVCLSSNKKDYDA 484
DB 475 LRKISKDAGMPIDQPCFCYKAGADSVEMFRLKNTYSG-LQLIIVIL--PGKTPYAE 532
QY 485 IKKYLCTDCPTPSQCVVARTLGKQQTWAIATKIALQNMCKMGE-----ELMRVDIPLK 538
DB 533 VKRIVGDTLLGMATQCVYKNAVK--TSPQTLISNCLKINVLGG--INNILVPHGRSAV 590
QY 539 LVMIYVIGDCYHDMTAG--RRSIAGFYASINEGNTWFSRCIFQDRGQELVDGLKVCLO 595
DB 591 PVILGADVTHP-PAGDGKPSITAAVGSMDAPSRXCATVRQRPQOEIIEDLSYWEELL 649
QY 596 ALRAMNSCNEMPSRIIVYDGVGDQGLKTLVNYEVPQFLDCKSIGRGVNPRLTVIVK 655
DB 650 LLIQFYSTRFKPRLIFRYDGVSEGOFRQVLYEELAIAREACISLEKDYQPIITIVVQ 709

```

```

QY 656 KRVNTRF-----AQSGLQNPPLPGTVIVDEVTRPEMYDFEIVSAVRSGSVSPETHNYI 711
DB 710 KRHTTRLEFCADRTERRVRSNGNIPAGTVDDTTHPEFDFYLCSHAGIGQTSRPSHYHL 769
QY 712 YNSGLKPEDEHIOQLTYLCHIIYNNPGVIRVAPCOYAHKLA- 754
DB 770 WDNCNCPADLQTLTYOLCHTYVRCRVSISIPAPAYAHVAPRAPHYLVKXEDSAGS 829
QY 755 -LVGQSIHREP 764
DB 830 HVSQSGNDRP 840

RESULT 13
12C4_HUMAN
ID 12C4_HUMAN STANDARD; PRT; 861 AA.
AC Q9HCK5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 4 (eif2c 4) (eif-2C 4)
DE (Argonaute 4).
GN EIF2C4 OR AGO4 OR KIAA1567.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res 7:273-281(2000)
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC -1- Initiation process (by similarity).
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 P1WI domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AB046787; BAB13393.1; ALT_INIT.
DB GenBank; F02170; PAZ.1.
DB MIM; 607356;
DB InterPro; IPR003100; PAZ.
DB InterPro; IPR003165; P1WI.
DB Pfam; PF02170; PAZ.1.
DB Pfam; PF02171; P1WI.1.
DB PROSITE; PS50821; PAZ.1.
DB PROSITE; PS50822; P1WI.1.
DB Initiation factor; Protein biosynthesis.
DB DOMAIN 225 338 PAZ.
DB FT DOMAIN 509 820 P1WI.
DB SEQUENCE 861 AA; 97096 MW; F236PF05047554C1 CRC64;

Query Match 12.1%; Score 495.5; DB 1; Length 861;
Best Local Similarity 24.0%; Pred. No. 7.7e-28; Indels 149; Gaps 35;
Matches 206; Conservative 149; Mismatches 353;
20 GSSGIIVRLSTHFLTRSPQWALYQYHIDVNLMEARLSALL-----FQHEDLIGK 73
22 GTVGKPIRLANHFQV-QIPKIDVYHYDDIDIKPRPRVREVVDTWVRFKQKQIFGR 80
74 CHAPDG-----TTLPLP---KRLQKVTVEFSKTNQGDVATITL----- 111

```

```

DB 81 QPGYDGRKNMTAPHLPIGRDRVDMETL-----LPBESKDQTFKYSVQWVSVDLLEA 136
QY 112 -----TNELPSPSPICLOEYNIIFRLLKIMNLQOIGSNVYVNDPIDP-SHRLVWPGF 166
DB 137 LAGHLEVPDSDVQALD-----VITREHLPSEKTYTVGASFSFSPGTYHPLDGGGEVWFGF 192
QY 167 TTSILQYENSIMLTCDVSHKVL-RSETVLDPMF-----NFYQTEBHKQEQY-SKEL 217
DB 193 HQSVRPAAMMMNLINVDGATAFYRAQPIEFMCVELDIQINIEQTKPLTDSQRYKFTXEI 252
QY 218 IGLVLTLYK---NNTYRVVDIDMDQNPSTF-----KKADSEVSFLYEYKQYNOEITD 270
DB 253 RGLKVEVTHGQMRKRCVNCVTRPASHQTFPLQLENGQAMECTVAQYFKQKSLQ--- 309
QY 271 LKQPVL-----VSQPRRRPGGTLFGPAMLIPELCTYLG-----LTDKRRNDFNMKD 319
DB 310 LKYPHLPCLQYQGRKH-----TYLPLEVQNIYAGQRCIKLTD---NOTSTM-- 354
QY 320 LAVHTRLTPRQPRQREVGRLDIYHKNDVQSELPDQGLSPDSNLSFSGRILQTEKHQ 379
DB 355 IKATARSAPD-RQEBISLTVKSNVSGGPPDYLEKEFGLVNEEMTELGRVLPAPMLQYG 413
QY 380 GKTPEY-NPQPADMSKETRGAPLISVKPLDNWMLITYRBYEANSILION-----LFKTP 434
DB 414 GRNKTVAIFNQGVM--DMRGKQFYAGIEIKYMAVACFAPQKQCRBDLKSFTDQLRISK 471
QY 435 AMGQMRK-----AIMEVDDRTFAYLRYLQOKVT-ADTQIVYCLLSNRKQYDAIKYL 489
DB 472 DAGMPDQGPCEFCRYAGADSEVEPMFKEL--KMYTVGLQTLIVTL-PEKTPYAEVKKVG 528
QY 490 CDDCCTPGCVVARTLGGQVPMATATKIALQNMCKMGEIMRVIDP-----LKLVMYI 542
DB 529 DTLGMAIQCVQKVVAV--TSPQLSLCLKINKKGG-INNVLPVHQRSVFGQPVIF 585
QY 543 VGIDCYHMDTAG--RSIAGFVASINEGMRWRSRCIFQDR-----GQELVUGL 589
DB 586 LGADVTLP-PAGDGKPKSIIAAVAGSMDSHPKRYCATVAVQTSROEISQELLYSGQVIDL 644
QY 590 KYCLQALIRAMNSCNEWMPSRITIVRDGVQQLKTLVNEVPCFLDCLKSIGRGYNRL 649
DB 645 TMMVBELIQFYKSTRFPRTRILTYRGVSEBQKQVAMPBLIARKKQISIEEDYRGI 704
QY 650 TVIVYKRVNTRF-----AQSGLQNPPLPGTVIVDEVTRPEMYDFEIVSAVRSGSVSP 705
DB 705 TVIVQKRHTRLFCADKTERVKGSGNVPAGTVDDTTHPEFDFYLCSHAGIGQTSRP 764
QY 706 TYNVITVNSGLKPEHIOQLTYLCHIIYNNPGVIRVAPCOYAHKLA- 754
DB 765 SHYQVLMDCNCPADLQTLTYOLCHTYVRCRVSISIPAPAYAHVAPRAPHYLVKXED 824
QY 755 -----LVGQSIHREP 764
DB 825 DSAEGSHVSGQSGNDRP 841

RESULT 14
AGOL_ARATH
ID AGOL_ARATH STANDARD; PRT; 997 AA.
AC Q9SJK3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute-like protein At2g27880.
GN AT2G27880 OR TIE2.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

```

RX MEDLINE=20083487; PubMed=10617197.
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.B., Feldblum T.V.,
 RA Buehl C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umeyam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasey T.H.,
 RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
 RA Nieman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.,
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana."
 RL Nature 402:761-768(1999).
 CC - SIMILARITY: Belongs to the argonaute family.
 CC - SIMILARITY: Contains 1 PAZ domain.
 CC - SIMILARITY: Contains 1 P1w1 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AC006929; AAD21514.1; -
 DR PIR, A84678; A84678.
 DR InterPro: IPR003100; PAZ.
 DR InterPro: IPR003165; P1w1.
 DR Pfam, PF02170; PAZ; 1.
 DR Pfam, PF02171; P1w1; 1.
 DR PROSITE, PS50821; PAZ; 1.
 DR PROSITE, PS50822; P1W1; 1.
 FT DOMAIN 359 471
 FT PAZ.
 FT FT 638 958
 FT P1W1.
 SQ SEQUENCE 997 AA, 111134 MW, DA93A1446C422F31 CRC64;
 Query Match 12.1%; Score 495.5; DB 1; Length 997;
 Best Local Similarity 24.1%; Pred. No. 9.5e-28;
 Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;
 QY 20 GSSGIIVRLSTNHPRLTSRQWALQYHNDVPLMEA-----RLRSALLFOHED--LIGK 73
 DB 154 GILGKWKVVRANHP--LVQVADADLVHYDVSNIPVISTKVRNWKLVKVKYKSHLGCK 212
 QY 74 CHAFD-----GTLFLPKRLQOKVTEVFSTKNGEDVRIITLITNELPSTPTCLOF 125
 DB 213 SPAYDORKSLYTAGPLPFDSEKFEVNLAEKRDGSSGKDRPKVAVKN---VTSTDLTQYL 269
 QY 126 VNIIRRR-----LLKIMNL-----QQIGRNVYNNPMDPIDISHR-----LVI 162
 DB 270 QQFLRKORREAPDYDTQVLDVLRDKPSNDVSVRSFPHSLGKDARDGEGELGDGIEY 329
 QY 163 MEGFTSLIQVENSIMLCTDVSHKVLRSSTVL--DFMFMFYHQTBEHK-----FOBQVSK 216
 DB 330 WNGVGSJLRLTQMGSLINIDVARSFYEPIVVTDTISKRLINRDLNRLPRLSDRLKVKV 389
 QY 217 LIGL-VLTLYN-NKTYRVDVDDMDONPKSTPKKADGVSSTLEYRQYQNEITDLKOP 274
 DB 390 LFTLTKRLAHMCTSAKISGISLPIRLRLPTLBDKSEKTVYQFAEKYRNVYQALP 449
 QY 275 VAVSQPKRRRGPGGLTGPAMLIPELC-----YLTGLTDRMRDFFVMKDLAVHTLT 327
 DB 450 ALQY-----GSTTRVYVLPMLCCQJDEGORTKRLNEKQ-----VTALLKATCOR 494
 QY 328 PQQRQREVRLLDYHKNDNVQRELDMGLSPDNLSSGGLQTE--KTHQGGKRTDY 385
 DB 495 PEPRENSIKLVKNNVNDLSKE--FGMSVTTQJLASLEAVLPPMLKYHDSSEKRYV 551
 QY 386 NPOFADWS-----KETGAPLLISVKPLDNLVLI--YTRRYEANSILQVLFVYTPAMQ 439
 DB 552 NPLRGKNNVLDKKNVNGAKVTS-----WTCVSFSTRIDRLGPPQFCGLIOMCYSKME 605
 QY 440 MRKA-----IMTEVDRTAEVRLQOKVTADTQIVVCLSSNRKDKDAIKKYLCTCP 494

Db 606 FKQPAIPFISCPPEIEEBLLDI--HKRPAQLQLIVL--PDVNGSYKIRICETELG 662
 QY 495 TFSQCVARTLQKQVTWMAIKIALQNNCKRGG-----ELMRVDLP---KLWIVGI 545
 Db 663 IVSQCCQPRQVKNL--KQYMEVVALKIVKTKGRTVLNDIAIRNIPILITRPTIMKA 720
 QY 546 DCHMDMTA--GRSINGFVASTN--EGMTWFSRCIFQDSGQELVDLAKYCLQALR-- 598
 Db 721 DVTHPQGEDSSSILAIVASMDWPEINKRGLVSAQAREELIQLYLVQDPQGLVH 780
 QY 599 -----AMNSCNEYSPRIIVYDGVQDQKTLVYVYVQFIDCLKS-IGRGYNEPL 649
 Db 781 SGILREHFLAFRATQIQIRIFRYDGVSEGFQSVLLHEMTALRKACNSJOENVVPPV 840
 QY 650 TVIVYKRVNTRFPFASG-----RLQNPLEGTVIDVEVTPREWDPIFVSAVRSASV 703
 Db 841 TFAIVQKRHRTRLPFQHGKRDNTDSGNIQETVVDIKICHPNEFDYLNHAGIGTS 900
 QY 704 SPTHYVVIYDNGSLKPDHICQLTYKLCIHYNNPQVIRVAPQYAKLAF 754
 Db 901 RPAHYHVLIDENGFTADQQLMTNNLCYTYARCTKSIVPXYAHLAF 951
 RESULT 15
 ID 12C2 HUMAN STANDARD; PRT; 851 AA.
 AC O9UKV8; Q8TC25; Q8WV58; Q9GID1;
 DT 16-OCT-2001 (Ref. 40, Created)
 DT 15-MAR-2004 (Ref. 43, Last sequence update)
 DT 15-MAR-2004 (Ref. 43, Last annotation update)
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
 GN EIF2C2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21912064; PubMed=11914277;
 RA Mourtelatos Z., Dostie J., Pauskhin S., Sharma A., Charroux B.,
 RA Abel L., Rappalber J., Mann M., Dreyfuss G.;
 RT "tRNPs: a novel class of ribonucleoproteins containing numerous
 RT microRNAs."
 RL Genes Dev. 16:720-728(2002).
 RN [2]
 RP SEQUENCE OF 231-851 FROM N.A.
 RX TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Altschuler S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang D., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,
 RA Brownstein M.J., Uedin T.B., Tochiyuki S., Carninci P., Prange C.,
 RA Bata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Melek J.A., Guarnarene P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickinson G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 267-851 FROM N.A.
 RA Kainline N., Chen X., Rolfe A., Halleck A., Hines L., Eisenstein S.,
 RA Koundinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin Y.,

RA Phelan M., Farmer A.;
 RT "Cloning of human full-length cDNAs in BD Creator(TM) system donor
 RT vector.";
 RI Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RX SEQUENCE OF 475-851 FROM N.A.
 RX MEDLINE=20005943; PubMed=10534406;
 RA Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
 RA Haasam S., Weitz S., Lichter P., Heltz P.U., von Knebel Doeberitz M.,
 RA Birner U.;
 RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
 RT genomic organization, localization to chromosomal bands 1p34-p35, and
 RT expression.";
 RL Genomics 61:210-218(1999).
 CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
 CC -1- initiation process.
 CC -1- SIMILARITY: Belongs to the argonaute family.
 CC -1- SIMILARITY: Contains 1 PAZ domain.
 CC -1- SIMILARITY: Contains 1 Pw1 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC
 DR EMBL; AY077717; AAL76093.1; -
 DR EMBL; BC007633; AAH07633.1; ALT INIT.
 DR EMBL; BC018727; AAH18727.1; ALT_INIT.
 DR EMBL; BT007229; AAB35893.1; -
 DR EMBL; AF121255; AAF13034.2; -
 DR GeneW; HGNC:3263; EIF2C2.
 DR GO; GO:0003743; F:translation initiation factor activity; NAS.
 DR GO; GO:0006412; P:protein biosynthesis; NAS.
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; Pw1.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Pw1; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; Pw1; 1.
 KM Initiation factor; Protein biosynthesis.
 FT DOMAIN 227 340
 FT 509 810 Pw1.
 FT 556 556 C-> W (IN REF. 4).
 FT 581 581 Q-> E (IN REF. 4).
 FT 609 609 S-> R (IN REF. 4).
 FT 629 629 E-> K (IN REF. 1).
 SQ SEQUENCE 851 AA; 96798 MW; 31A641ABC7D1DDB4 CRC64;
 Query Match 11.9%; Score 487; DB 1; Length 851;
 Best Local Similarity 24.0%; Pred. No. 3,1e-27;
 Matches 192; Conservative 146; Mismatches 341; Indels 120; Gaps 32;
 QY 39 PGMALYQYHIDYNDPLMEARLRSLAL--LFOH--EDLIG-KCTAFDG-----TLPFPR 87
 DB 42 PKIDYHYELDIEKPKERRVRREIYEMVQHFKTQIFGRKPVDFGRKNLYTAMPPIG 101
 QY 88 LQOKVTEVSKTRNGED--VRIT-----TLNMLPPTSPCLQFYNIFRR 132
 DB 102 -RDKVELEVTLPGRGKDRIFKYSIKWVSCVSLQALHDALSGRLPSVFETIQALDVMRH 160
 QY 133 LKIKMNLQOIGRNTYNNPDIPIDIP-SRLVIWPGFTTSILOYENSIMLCTDVSHKVL-RS 190
 DB 161 -LPSWKRYTPVGRSFTASBGCSNPLGGREYVFGFHQSVPRLMKMLNIDVSATAFYKA 219
 QY 191 ETVIDF---MENFYQTEEHK-----FOGVSKELIGLVVLTXY---NNKTYRVDIDWD 239
 DB 220 QPVTIEVCEVDLFSIEQKPLDSDQVYKFTKEIKGLKLVITHCGQMKRYKRYCANTRR 279
 QY 240 QNPKSTFKKADG---EVSLEYRKQYNQETIDLKQPVL-VSQPKRRRGPGGTLPGA 294

DB 280 PASHQTFPLQSGQVETCAQYFKORHKLVIRYPHLPCLQVQEOQH-----T 329
 QY 295 MLIPELCYLTG-----LTDKMRNDFVMKDLAVHTRLEPCQROREVGRLIDYHKNDN 347
 DB 330 YLPLEVCNIYAGQCIRKLTLD---NQTSTM--IKATRSAPD-RQEIISKLMR--SASFN 381
 QY 348 VQREIRDMGLSPDSNLSFSGRILLQTEKIHOGKTFPDYNPQFADMSKETRAPLISYKPL 407
 DB 382 TDPYVREFGINVKDEMDDVTGRVLQPELIVGR-----NKALATPVQGVWDM 429
 QY 408 DN-----MLIYTRNVEAA-----NSLIQNLFKVTPAMQMRK-----AIMEV 448
 DB 430 RKKQPHTEIKRWAIACFAPQROCTEVLHLSFTEQLRKISRDAGMPLQGPFCCKVAQG 489
 QY 449 DDTREAYLRVLIQKVTADTQIVCLSSNRKQKTDALIKYLCCTCPSPSGCVVARTLGKQ 508
 DB 490 ADSVPMFRHH-KMYAGLQLVVIL-FGKTPVVAEVRVGDVTLGNATQCVQKNV--Q 545
 QY 509 QTVMAIAKIALQNMCKMG-----ELMRVDIPKLVMIVGIDCYHDMTAG--RSIA 559
 DB 546 RTTPTLNLCKIKIVKLGVMNNILLPGREPPVQQPVIFLGADVTHP-PAGDCKKSSIA 604
 QY 560 GFVASINSGMTRWFSRCIFDQGGELYDGLKVCLOALPAMNSCNEYMPSRIIYRDVG 619
 DB 605 AVVGSMDAHPNRKCATVAVQHQEIIQDLAAVRELLICFKSSTRFKETRIIFYRDGVS 664
 QY 620 DQGLKTVNYVYVPOFLDKSIGGYNRLVVIYKKRVNTRFFA---QSGRLQNP 675
 DB 665 EQFOQVLAHHEILAIREACIKLEKYOQGITFIYQKSHHRLCTDGNERNVSGSINPA 724
 QY 676 GTVIDEYTRBEMVDFFIVSQAVRSGSVSPHYNYIYDNGSLKPDHIIQLTYKLCIIYNN 735
 DB 725 GTTVDTKITHPTBEPFLCSHAGIQGTSRPSHHVLMWDMRFSDELDILTYQLCHTYVR 784
 QY 736 WEGTVRVAPPCOYAKLAF 754
 DB 785 CTRSVSIPAPAYVAHLVAF 803

Search completed: May 5, 2004, 15:13:09
 Job time : 21 secs

Wed May 12 09:50:22 2004

us-10-043-774b-2.rpr

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:10:30 ; Search time 20 Seconds

(without alignments)
3727.422 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVNTRQNDHVESKTKG.....VGOSIHREPNLSLNRLYYL 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1295.5	31.6	824	2	T20351
2	1162	28.3	722	2	T30995
3	557.5	13.6	834	2	T41568
4	508.5	12.4	1014	2	H86438
5	495.5	12.3	891	2	T32079
6	488.5	12.2	1000	2	T23391
7	493.5	12.1	997	2	A84678
8	492	12.0	988	2	T52134
9	488.5	11.9	930	2	A84668
10	482	11.8	813	2	JC6569
11	434	10.6	897	2	T01113
12	429.5	10.5	1032	2	T23164
13	429.5	10.5	1035	2	T23165
14	421	10.3	958	2	S41013
15	421	10.3	1040	2	D85668
16	421	10.3	1040	2	D85668
17	343.5	8.4	1139	2	T33510
18	316.5	7.7	905	2	T23510
19	283	6.9	892	2	T15195
20	280.5	6.8	945	2	T23965
21	253	6.2	965	2	T22933
22	252	6.1	1057	2	T16676
23	248.5	6.1	1057	2	T19324
24	240	5.9	1030	2	T16114
25	238	5.8	990	2	F88925
26	230.5	5.6	697	2	T15179
27	225	5.5	939	2	T18974
28	223	5.4	324	2	G87774

30	219.5	5.4	674	2	T19268	hypothetical prote
31	219	5.3	363	2	T34486	hypothetical prote
32	200	4.9	567	2	T34339	hypothetical prote
33	189.5	4.6	706	2	H70425	conserved hypothet
34	168	4.1	944	2	T18911	hypothetical prote
35	145	3.5	713	2	H84464	hypothetical prote
36	129	3.1	185	2	T27784	hypothetical prote
37	129	3.1	709	2	G96610	probable disease r
38	123.5	3.0	2035	2	AC0233	yeastinabactin bio
39	123.5	3.0	2041	2	T17439	peptide synthetase
40	120	2.9	1261	2	S75130	sensory transducti
41	118	2.9	1120	2	H71664	transcription-repa
42	114.5	2.8	474	2	A64691	type III restricti
43	114	2.8	2035	2	A48654	probable polyketid
44	113.5	2.8	560	2	C38604	poly(3-hydroxyalka
45	113	2.8	949	2	G84486	probable plasma me

ALIGNMENTS

RESULT 1

T20351
hypothetical protein D2030.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T20351

R/Wilkinson, J.
submitted to the EMBL Data Library, June 1996

A/Accession: T20351

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-824 <WTL>

A/Cross-references: EMBL:273906; PIDN:CAA98113.1; GSPDB:GN00019; CESP: D2030.6

A/Experimental source: clone D2030

C/Genetics:

A/Map position: 1

A/Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3

Query Match
Best Local Similarity 31.6%; Score 1295.5; DB 2; Length 824;

Matches 278; Conservative 169; Mismatches 302; Indels 47; Gaps 15;

QY	16	BSKTGSSGIIIVRLSTNHFRL-TSRPQALQYHIDVPLMEARRLRGALDFQHE--DLIG 72	
DB	40	QSKTSSGQPKCFANPIENTQDSYIYQHYVEFPYDSKANKRKLDDNNVDEIG 99	
QY	73	KCHAFDGTILFLPRLQQYVEFSKTRNGEDVATITLTNELPTSPFCLQFYNIIFR 132	
DB	100	HHFVFDGMLLYLKXEMFQNMQIEVQHPIDRLICIRKQNRFLVDPPOTINIFNTIR 159	
QY	133	LIKTNMCOGRNRYNNDPIDPSHSLVIMPGFTSILQYENSIMCTGVSHKVSSET 192	
DB	160	SPDALQTLQGRNRYNNDPIDPSHSLVIMPGFTSILQYENSIMCTGVSHKVSSET 219	
QY	193	VLDPMFNFYH-----QTEHKKPEQYSKELIGLVLTKNYNNKYRVDIDIDQPKSTF 246	
DB	220	---MYLIFKHELRSCQNNQORVQKKNEMVGGTITITRYNNKLRFTYRDIYSISPSSEF 275	
QY	247	KKADGSEVSFLFYRKYQNGEITDLPVLSQPKRRGCGTLPGSAMLIPELCYTLGL 306	
DB	276	VR--DGQSITLTKYKQNGEITDLPVLSQPKRRGCGTLPGSAMLIPELCYTLGL 332	
QY	307	TDKRNDFNWKDLAVETRLTPEQORREVGRLIDYIHKNDNVQRELRDWSLSPDSNLSF 366	
DB	333	TDEKRDQFQWKEIAKNTIMSPOQRLVESRLIYDLSKNEKVMGCFKMGISLQDLANV 392	
QY	367	SGRILOTEKIHGGKTFDYNQPADMSKETRGAPLISVKPLDWLLTYRNTYA--ANS 424	
DB	393	QARVLKSEPL-QGKKT--YEGKQAEWARGVGECIGYRGSNMNTWIVIGPSGNSGLLSQK 449	

425 LIQNLFFKVTTPAMGKQKRAIMIEV---DPRTEAYLFLVQOKTADTQIVCLLSNRK 479
 Db 450 FLEEARRLGKILIQVQLGSPMCVPIRGISIPNDYLEGVKAIQVQGEDIHMLVWLADNKK 509
 Qy 480 DKYDAIKKYLCTDCTPSCQCVARLT-----GKQCTVMAIATKIALQNNCKKMGELMR 532
 Db 510 TRYBSLKKFLVCECIPQCYNRLTLAKSKDGGENKRLSVLKIYQMLCKTGALNKK 569
 Qy 533 VDIPLKLVYIGIDCYHMTAGRSIAGFVASINEGTRMFSRCIFODRGOGLVDGLKVC 592
 Db 570 VNIPLKNTMIVGYDYLDHSTLKGKTVGACVSTSTNDPQFYSQTRPHENPTQLGNLTHF 629
 Qy 593 LQALIRANNSCNEY-NPSRIIYVRDVGDKQKLTNNVEVPOFLDCLSKISIR----- 643
 Db 630 VRKALKQYDSNDQTLPSRLILYRDGADGQIPYIKNTEVGLVNDACDCAVTDKAAELSNK 689
 Qy 644 -GYNRLTVIVYVKKVNTFRFAQSGRLQ---NPLPGTIVDEVTREPMYDFIVSOAVR 699
 Db 650 VOEKIKLAFIIVTKGVNNRILKQ--GSSIDNAINPQPGTVDDTVTRPERMDFYLVQFVN 748
 Qy 700 SSVSPHTNNVYDNSGLKPDHIGRLTYKLCIYNNMPGVLRVAPQCYAHKLAFLVQGS 759
 Db 749 QCTVPSVNIHDDTIDGPDGQQLAFKLCILYNNMGTVRVAPQCYAHKLAFLVQGS 808
 Qy 760 IHREPNLSLSNRLLYL 775
 Db 809 LHDDANGCLRDLPFL 824

RESULT 2
 T3095
 Hypothetical protein C01G5.2 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
 R/Bradshaw, H.; Steellies, L.
 Submitted to the EMBL Data Library, August 1999
 A/Description: The sequence of C. elegans cosmid C01G5.
 A/Reference number: Z20956
 A/Accession: T30955
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-722 <BRA>
 A/Cross-references: EMBL:U50068; PIDN:AAB37734.1
 A/Experimental source: strain Bristol N2
 C/Genetics:
 A/Map position: IV
 A/Intons: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
 A/Note: C01G5.2

Query Match 28.3%; Score 1162; DB 2; Length 722;
 Best local similarity 33.6%; Pred. No. 2.7e-76;
 Matches 256; Conservative 154; Mismatches 279; Indels 74; Gaps 14;

Qy 41 WALYQYHIDYVPLMARRLSALYFOHEDL--IGKCHAFDGTILFPLKRLQOKTAVFSK 98
 Db 6 YSIYQYHVEPPTVDSKATREMLAFQPSVYVEIGHFVFDMLILYKEFMQNMIVQH 65
 Qy 99 TRNGEDVAITTLTINELPPTSCLQFYNIIFRLKLKTMNLOIGRNYNPNDDIDIPSH 158
 Db 66 PNDNSLICIRPKTKRFLVDDPQTINIFNTIIRSFDMKLTQIGRNYFD----- 115
 Qy 159 RLVMPGFTTSLQYENSIMLCTDVSHKVLSEYLDPMFNFYQTEBHKQEOVSXELI 218
 Db 116 -----WNSBAL-----RKELNS-----CONRQKQESKNEYG 145
 Qy 219 GLVLTJKNKTYRVDDIDWDONPKSTFKKADGSEVSFLEYRQYNOETIDLKQVLYS 278
 Db 146 GSTIITRYNNKILHRTFLRDLNEITPLSKFOK--DGEQIILKEFYKQYDIDIDDEQFIIS 204
 Qy 279 QPKRRRGGGLTGPAMLIPELCYLTGLDKKRNFNVMKDLAVHTRLTPEQORQREVRL 338
 Db 205 ESKPKQ--FGEPQVNYIPELCTPGLTDEKRDFOKMKELNHTTSSFOQRDLTRKL 262

Qy 339 IDYIHKNDVQRELRDWSLSPDSNLISFSGRLIQTEKHQGGKTDYNDQPADWSKETNG 398
 Db 263 ITLISQNTMMECFQWYGISLQODLANVQARLYKSEPL-QGR--QYEGKQEMARVGE 319
 Qy 399 APLISVKPLDNNLLIYTRNRYE--ANSILQVLFKYTPAMGQKRAIMIEV-----DDR 451
 Db 320 CGIYRSNNMTMIVIGPGSGNSGLAQKTAARNLGRQLQVQLGEPKCVKINGISPNY 379
 Qy 452 TEAYLVLQOKTADTQIVCLLSNRKQKYPALIKYLTCTDCTPSCQCVARLT----- 505
 Db 380 LELIKAIKSDVGEIEHMLVWLADNKKTRYLSLKKYLCECPPIPQCYNRLTLAKSKD 439
 Qy 506 -GKQCTVMAIATKIALQNNCKMGELMRVDIPLKVMIVGIDCYHMTAGRSIAGFVAS 564
 Db 440 GGENKRLGSIYLVKIYQMLCKTGALNKNITELKSMIVGYLDYHDSITLKGTVGACVST 499
 Qy 565 INEGTRMFSRCIFODRGOGLVDGLKVCLOALIRANNSCNEY-NPSRIIYVRDVGDDGL 623
 Db 500 TSNDFQFYSQTRPHENPTQLGNLTHFYRKSLKQYDNNDKTLPSRLILYRDGADGQI 559
 Qy 624 KILVTEVQFLDCLSKISIR-----GYNRLTVIVYVKKVNTFRFAQ--SGGRLQ 672
 Db 560 PYIKNTEVGLVNDACDCAVTDKAAELSNKVQEKIKLAFIIVTRVNNRILKQSSSKSAIN 619
 Qy 673 PLPGTIVDEVTREPMYDFIVSOAVRSVSPHTNNVYDNSGLKPDHIGRLTYKLCI 732
 Db 620 PQPGTVDDTVTRPERMDFYLVQFVNOGTVPVSNIIHDDTGLGPDGHQQLAFKLCIL 679
 Qy 733 YNNMPGVIRVAPQCYAHKLAFLVQGSITREPNLSLSNRLLYL 775
 Db 680 YNNMGTVRVAPQCYAHKLAFLVQGSILHDDANGYLRLDLPFL 722

RESULT 3
 T41568
 argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C/Accession: T41568
 R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
 Submitted to the EMBL Data Library, May 1998
 A/Reference number: Z21991
 A/Accession: T41568
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-834 <WCO>
 A/Cross-references: EMBL:AL023705; PIDN:CA119275.1; GSPDB:GN00068; SPDB:SPCC736.11
 A/Experimental source: strain 972h-; cosmid c736
 C/Genetics:
 A/Gene: SPDB:SPCC736.11
 A/Map position: 3
 A/Intons: 43/3
 C/Superfamily: rabbit translation initiation factor eIF-2C

Query Match 13.6%; Score 557.5; DB 2; Length 834;
 Best local similarity 22.9%; Pred. No. 3.1e-32;
 Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

Qy 20 GSSGIIVRLSTYHFRITSPQALVQYHIDYVPLMEARRLSALLFOHEDL----- 70
 Db 16 GGLGKQITLKANFFQIISLPNETIQYHIVDGSVRPKSQQLIWSKENVQYFGSSWY 75
 Qy 71 -----IGKCHAFDGTILF-----LPKRLQ--QKYTEV-----FSKTRNGED 104
 Db 76 NSVYDGRSMCKSGDLADGTIKNIGSSESHPEISIOKSKINLHTLSQVNSKSYSD 135
 Qy 105 VRTITLTINELPPTSCLQFYNIIFR-----LKLKMLQOIGRNYNPNDDIDIPSHR 159
 Db 136 -----FOVLSSIMFIDLILKKRPSSETLRFQFMSFPTGNGVSLGGVB----- 178
 Qy 160 LVIMPFTTSLQYENSIMLCTDVSHKVL-RSEYVLDPM----- 197

Db 179 --AWGFGVQSRPNQFMSVAVNDISSAFWENDSLQILMEYTDSCNVDRDITRDLKRLS 236
 198 --FNFYHOTEHKEHKEQVSKELIGLVLTNNKTYRVDIDWDONPKSTF--KKADGSE- 253
 Db 237 RKFRLKATYTCQRRNN-----VGTDLANRVYSIEGSSKASDSEFVRLNGEQ 285
 Qy 254 -VSFLFYRKQYNOGIIITLQKQPVLYSQRRRGPGTLPGRAMILPELCYLT---GLTDK 309
 Db 286 KISVAEYELNNVRLQYPNLPCTIL-----VKNQAMLPFECFVVKQKQRTAK 333
 Qy 310 MENDFNWMDLAVHTRLTPEQRFQREVGRLLIDYHKNQ--NVQRELRDWGLSPDSNLSFSG 368
 Db 334 LNSD-----QTANMRFVAVQRPFRVQIIDFVHQMDWDTDPYLTQYGMKIQKQMLEVPA 388
 Qy 369 RILQTEKHOGKRTDYNPQRPADMSKETRGAPLISV--KPLDNTMLI---YTRR-NYEAA 422
 Db 389 RLETPSIRYGGDCTE-RPVSGRNN--LRGKRFDPFPAPFIRSMVAVKCFSTIRLPMKGI 445
 Qy 423 NSLIQNLKFTVPAMQO--MRKAIMIEVDDR--TEAVLRVLQOKV---TADTOIVCLL 474
 Db 446 ENPLQTYVQTLTSLGINFVWKRPVLVADINGSVELCITLYKKAQYGNAPPDYLPETL 505
 Qy 475 SSNRKQTDALIKYLCCTDPPSQCVARTGKQQTNAIKTALQWCKMG---ELW 531
 Db 506 DKNSPEPVSIGKRVGNITMLGVPSQCAISKHI--LQSKPOYCANLGMKINVKVGINGSCLI 563
 Qy 532 RVDILKLV--MIVGIDCH--DMTAGRSIAGFVASINSGMTBMFSRCIFDRQGLVDG 588
 Db 564 PKNPGLNVPITLLGDVYHGVGATGVSIASIVASVDLNGCKRTAVSRQPRHDEVIG 623
 Qy 589 LKVCLOALFVANSNCEWYPSRIIYRQGVGGQKTLVYEVPPFLDCLKSIGRGNPR 648
 Db 624 MDIVVILLQFRAMTKQOPRIIYFRDQTSQGLVYNDELQIKKACSLSPKYNPK 683
 Qy 649 LTVIYVKKRVNTRFRAQS---GGRLONPFGTVIDEYTRPBNMFFIVSQANSGSASP 705
 Db 684 IIVCTQKKHARFPTKNSDSDNGNPLPGTILKHYHVPYQYFYIISHPSLQGSVP 743
 Qy 706 THYNYIYDNGSLKPHIQTLTYKLCHIIYMPGVIRVAPCOYAKHAFI----- 755
 Db 744 VAYTTLHBDICMPPOQFQTLCTNLCYVBARASVILVPVYALVSNLAKYQDVTAD 803
 Qy 756 -----VQGSIRHEPVLISNKL 772
 Db 804 TEVETSEASMDQEV--KPLALISKL 827
 RESULT 4
 H86438
 Protein T19E23.7 [imported] - Arabidopsis thaliana
 C1Species: Arabidopsis thaliana (mouse-ear cress)
 C1Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C1Accession: H86438
 R1Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Matzila, I.
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A1Author: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A1Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A1Reference number: A86141; MIMD:21016719; PMID:11130712
 A1Accession: H86438
 A1Status: preliminary
 A1Molecule type: DNA
 A1Residues: 1-1014 <STO>
 C1Genetic: GB:AE005172; NID:g6692120; PIDN:AAF4585.1; GSPDB:GN00141
 A1Gene: T19E23.7
 A1Map position: 1

Query Match 12.4%; Score 508.5; DB 2; Length 1014;
 Best Local Similarity 23.9%; Pred. No. 1.5e-28;
 Matches 209; Conservative 151; Mismatches 345; Indels 171; Gaps 40;
 Qy 5 VNTKQMLDHYKSKTSSGII-----VRLSTNHLRSLRPMALYQYHIDN---PLMEAR 57
 Db 153 VATSDEKPEKRPDRG--GVAVRNVNLVYHKKVFNESVIRHYDVEIKGEIPTRKYS 210
 Qy 58 RLASALLPQH-----EDLIGKHAQD-----TILFLKRLQOKTEVSKTRNGEDVR 106
 Db 211 RFLANRQKRVFTDNDEPPLANTAYDQGNIFSAVELPTG--SYKVEYKTEMRGRST 269
 Qy 107 IITTLNLEPPTSPTCLQFYNIIFRLIKIMLQO--IGRNYNPNDP-----DIP 156
 Db 270 FTIKQVN-----VLKIGDLKEYMTGRSSFNPRDVLQGMVWKEHP 310
 Qy 157 S-----HRLVMPGFSTSLQYENSIMLCTDYSHKYLR--SETV 193
 Db 311 SKCMITVGSFTRTEPDEDFRFGVIAKGRHILKPTAQSLCLDYSVLAFAKMSV 370
 Qy 194 LDMFNFYHOTEHKEHKEQ--VSKELIGLVY-----LTNNKTYRVDID 237
 Db 371 IEYLKLYFNMSDMRQRRRVEEELIGKATVNHKKQKGLTIVGLSMQTKDIFDLID 430
 Qy 238 WDQN--PKSTFKADGSEVSFLFYRKQYNOGIIITLQKQPVLYSQRRRGPGTLPGRAM 295
 Db 431 QEGNEPPEPKT-----SIVEYFRIKYGRHIVHKDIPCLDGKNGRQ----- 471
 Qy 296 LIP-ELCYTLGLDKKRNDFNWKDLAVHTR-----LTPEQRFQREVGRLLIDYHKNQVQR 350
 Db 472 FVMEFCDL--VGGQIYPKDNLKDSALMLKLSLVNPOQRGNIDMKR--ARNGSGGE 528
 Qy 351 ELRDWGLSPDSNLSFSGRILOTE--KHQGGTFDYNP--OPADMSKETRGAPLISVK 405
 Db 529 IINFGKLVDTNMTPEAGRVLKAPSLKAEGRVVEEPNPRQNNQNMKKQVTRGSI- 587
 Qy 406 PLDNTMLI--YRRRYEANSILQNLKFTVPAMQOYAKHAFI-----DTEA 454
 Db 588 -YRNMVAVLDTASERFNKPNDFVNDLIDRCMLQGMQEMAPYIKYSRMTLSNGNAIE 646
 Qy 455 YLRVLQOKV-----ADTOIVCLLSNRKDYAKIKYLCCTDPPSQCVNA--RTLQK 507
 Db 647 LRSVIDEASRKHGARGPTVLQAM--SRKODGKTLKIMIEFKLGLVTOCFGLGPAITKG 705
 Qy 508 QQYVMAIATKIALQWCKMG---ELKRVDPILKL--VNVIGIDCHMDTAGRR--SIA 559
 Db 706 DQ-----YRANLALKNNAKVGSSNVELDPTFSFKKEDVMTFQADVNHPARDKMSIV 761
 Qy 560 GFVASIN-EGMTWFSRCIFQDRGQELVLDL-KVCLQALFVANSNCEWYPSRIIYVRDG 617
 Db 762 AVYGTLMWPMANVAAVIAQHPKKEIQQFQDACE-LVKAHVQATGKAPNKIVIPRDG 820
 Qy 618 VGGQKTLVNYEVPQFLDCLKSIGR-GYNRLTVYVKKRVNTRFPA--QGGGRQNP 673
 Db 821 VSDAQPFMVLNVEI--LDVKLFKKNGXNPKTIVIAQRHDTFRFPATNNDOSDKGNV 877
 Qy 674 LPGVVIDEYTRPBNMFFIVSQAVRSQVSPTHYNYIYDNGSLKPHIQTLTYKLCHY 733
 Db 878 PSGTIVVTKYHPEYFYLCSHHGIGTGKPHYIYLMBELFTSQOVQKLFEMQFTF 937
 Qy 734 YNMPGVIRVAPCOYAKHAFIYVQSGIH-----RPN 765
 Db 938 TRCTKPVSLVPVYADMAVFR-GRWYTHASSRBN 972
 RESULT 5
 T32079
 Hypoetical protein T07D3.7 - Caenorhabditis elegans
 C1Species: Caenorhabditis elegans
 C1Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C1Accession: T32079
 R1Author: B.; Wamsley, P.; Tyman, B.
 submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid T07D3.

A:Reference number: Z21121

A:Accession: T32079

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-891 <LAM>

A:Cross-references: EMBL:AF016682; PIDN:AB66187.1; GSPDB:GN00020; CESP:T07D3.7

A:Experimental source: strain Bristol N2; clone T07D3

A:Gene: CESP:T07D3.7

A:Map position: 2

A:Insertions: 34/2; 206/1; 560/1; 879/3

A:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

QY 748 YAHKLAF 754
DB 835 YAHLYAF 841

RESULT 6
T22391
hypothetical protein F48F7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T22391; T22324

R:Colles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19559

A:Accession: T22391

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1000 <MTL>

A:Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone F48F7

R:Colles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19712

A:Accession: T22324

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1000 <MT2>

A:Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone K02B9

C:Genetics:

A:Gene: CESP:F48F7.1

A:Map position: X

A:Insertions: 70/3; 128/2; 185/2; 673/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

```

Db 559 PKLYGKTRATYALPNOGV--DMEGKQFHTGIDIVRWALACPAQOQKVENDMKMTNQ 616
Qy 429 LEKYPAMQOM-----RRAMIEVDDRTYALRVLOQKVTADTOIVVCLLSNRKDK 481
Db 617 LGRSNAGMIVGNPCCKYAVGE---QVEPMFKLKQYSG-IGLVVVL-PRGTPV 671
Qy 482 YDAIKYICTCPPTPSQCVAVATLGKQOTVMAIKTALQNNCKMG-----ELMRVDI 535
Db 672 YAEVRKVDIVLGATQCVQAK--NAITTPQTLISNCLKNVGLGQVNSILLPNVPRI 729
Qy 536 PLKVMIVGIDCYHDMTAGRR--SIAGFVASINEGMTWFSRCIFODRGQELVDGLKYL 593
Db 730 FNEPIFFGCDITHPPADSKRPSIAAVGSMADHPSYATVAVQCHRGSLISDLVYV 789
Qy 594 QALPAAWNSCNEWMSRIIVYRDGVGDQQLTVNYEVPQFLDCLKSGIRGYNPRLTVIV 653
Db 790 RELVQFRNTRFKPARIVVYRDGVSEGFNVLIQYELRAIREACMMLERGYQGITFIA 849
Qy 654 VKKRVNTRFEA-----QSGGRQLNPLPGTVIDVETVRPEWDFEIVSQAVRSGSVPTPHN 709
Db 850 VQKRHTLPAVDKQDVQKAVNIPPGITVDVGIHTEFDFYLCSHAGIGTISRPSHH 909
Qy 710 VIYDNGSKPDHIQRLTYKLCHIIYNNPQVIRVPAPCOYAHKLA 754
Db 910 VLMDNNLTADLQQLTYQMCHTVYRCSRVSIPAPAYALVAF 954

```

RESULT 7

```

A:Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84678
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.B.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617137
A:Accession: A84678
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-997 <STO>
A:Cross-references: GB:AE002093; NID:g4510428; PIDN:AD21514.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g27860
A:Map position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C

```

Query Match

```

Best Local Similarity 12.1%; Score 495.5; DB 2; Length 997;
Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;

```

```

Qy 20 GSSGIIVLSTNHFLTSRPMALQYHIDVNPMEK-----RLRSALLPQHEP-LIGK 73
Db 154 GTLGKVMVNRNH-LVQVADRDLHYDVSINPEVISKTNNVKNKLVKNYKDSHLAGK 212
Qy 74 CHAPD-----GTLFLPKRLQOKVTEFESKTRNGEDVRTITLTNLEPPTSPTCLQF 125
Db 213 SPAYGRKSLYTAGLPFDSKSEFVNLAEKRADSGSKDRPKVAVK---VTSDIQL 269
Qy 126 YNIIIR-----LTKTML-----QOIGRNYNPNPDIISHR-----LVI 162
Db 270 QOFLRKOREAPYDTIQVLDVLRDKPSNDVSVGRSPFHTSLGKADBGELGADGLEY 329
Qy 163 WGGFTSLIQVENSIMLCTDVSHKVLRSFTVL-DMFNFYHQTEBK-----FQEQVKE 216
Db 330 WGGYFQSLALQMGSLINDVARSFYERIVVTDPIFSKFLNIRDNLNRPDSRLKVKV 389
Qy 217 LIGL-VLTKY-NKTYRVDIDWDONPKSTKADGSEVSLFYRKQYNOETIDLKOP 274

```

```

Qy 275 VLVSQPKRRRGGGILPGPAMLIPELC-----VLTGLTDKMRNDFNMKDLAVHTRLT 327
Db 450 AIQT-----GSTRFVYVPMELCOIDEGQYTKRINKEQ-----VTALLKATCR 494
Qy 328 PEQQRREVRLIDYHKNDNVORELDMGLSPDSNLSFGRILOTE--KHQGGRTPDY 385
Db 495 PDRRENSIKLVYKNNYNDLSKE--FGMSYTTQASIEAVLPPMLKYHDSGKRYV 551
Qy 386 NPQPADMS-----KETRGAPLISVKPLDMLL--YTRRYEAANSIQVLFVTPAMNQ 439
Db 552 NPELGQNMNIDKQNVGAKVTS-----WCVSFSITRDRGLPQEFCKQIGKCVSKME 605
Qy 440 MKKA-----IMIEVDRTYAVLRVLOQKVTADTOIVVCLLSNRKDKYDAIKYICTCP 494
Db 606 FKQPAIPFISCPREHIEBALDI--HKAPGLQLIYVL-PDYGSYKIKRICEBELG 662
Qy 495 TPBQCVVAATLGKQOTVMAIKTALQNNCKMG-----ELMRVDIPL--KLWIVGI 545
Db 663 IVSQCCQPRQVNNKN--KQYMERVALKIKVKTGRTVINDAIRNIPITDRPTIMGA 720
Qy 546 DCHDMTA--GRRSIAGFVASIN-EGMTWFSRCIFODRGQELVDGLKVCLOALR---- 598
Db 721 DVTHPQPEDSSEPSIAVVAWSMDEPKIKRGLVSAQAREBELIQLYLVQDPQKGLVH 780
Qy 599 -----AMNSCNEWMSRIIVYRDGVGDQQLTVNYEVPQFLDCLKSGIRGYNPRL 649
Db 781 SGLIREHPIAFRATQIPIQRIIFYRDGVSEGFQVLHEMTAIRKACNSIQENYVPRV 840
Qy 650 TVLVYKKRVNTRFEASG-----RLQPLPGTVIDVETVRPEWDFEIVSQAVRSGSV 703
Db 841 TFIYVQKRHTLRFPQHARDMTDSGNIQPTVVDYTIKCHNEDFLNSAIGQGIS 900
Qy 704 SPHYNVVIYDNGSLKPDHIQRLTYKLCHIIYNNPQVIRVPAPCOYAHKLA 754
Db 901 RPAHYVLVDENGFMTDQQLMTNLCYTAGCTSVSLVPAYVHHLAF 951

```

RESULT 8

```

T52134
Zwille protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52134
R:Laux, T.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z25975
A:Accession: T52134
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-988 <LNU>
A:Cross-references: EMBL:AJ223508; PIDN:CAA11429.1
C:Genetics:
A:Gene: zwille
A:Map position: 5
C:Superfamily: rabbit translation initiation factor eIF-2C

```

Query Match

```

Best Local Similarity 12.0%; Score 492; DB 2; Length 988;
Matches 204; Conservative 133; Mismatches 397; Indels 132; Gaps 32;

```

```

Qy 5 VNTRONLDHYKESKSSGIIIVLSTNHFLTSRPMALQYHIDVNPMEARLSAL- 63
Db 120 MGRKNSNFAPRGFGTLGKCYKANHF-LADLPFKDLNQDVYITPEVSSKSVRAII 178
Qy 64 -----LFQHEDLIGKCHAPD-----GTLFLPKRLQOKVTE-----VSKTRNGEDV 106
Db 179 AEIVRLYKESDLRRRLPAYGRKSLYTAGELPFTWKEFSYKIVDEDDGILNGRKRERYK 238
Qy 107 ITTT-----LTNLEPPTSPTCLQFNIIFRLIKTMINLQOIGRNYNPNPDI 153
Db 239 VALKTVARAMNHHLGELTAKRADCPQEAQIIDIYLR--LSYKRPQVGRSPFSFD--I 295

```

QY 154 DIP---SHRLVWPGFTTSIIQYENSIMLCTDV-SHKVRSBETVLDMPNFYHTEHRK- 208
 DB 296 KTPORLGBLSWCGFQSIIRPTQMGSLNIDMSAAFIPELPIEFAVQLLGKDVLSKP 355
 QY 209 ----FOEQSKELIGLVLTAKN---NKTFRVDIDMDQNKSTFKKADSGEV-SPLEY 260
 DB 356 LSDSRVAVIKKGLRGVKEVTHRANVRKAVAGLTTPTELMFEPDENOTKMSVLEYF 415
 QY 261 RKQINQETIDLKQPVY-VSQPRRRGPGTLPGRAMLIPE-LCYLTGLTDMKRDVPMK 318
 DB 416 QEMVGTITQHTHLPCLQVGNQK---ASYLPMKCKVEGQRYTKALNEKQ----- 463
 QY 319 DLAVITRLTPEORREVRLLIDYI---HKQDNVQRELRDWSLSPDSNLSFGRIQTE 374
 DB 464 ----ITALKTYCQABEGQNDILTYQHNAYDQDPYAKERGMNISEKLASVEARILPAP 519
 QY 375 --KIHQGGKTPDYNOFADWSKETRGADLISVKPLDMLIYTRNTE--AANSLIONLF 430
 DB 520 WLYKHENGKCKDCLPQVQGMNMNKK--KQINGMTVSRAWACVPSRSYQENVARGFCNELG 577
 QY 431 KTPAMGMQMRKAMIEV---DQRTS---AYLRVLQCKVTADTOIVCLISNKKDK 481
 DB 578 QMCVSGMENEPEPIFYISARPDQVEKALNHYHTSMKTKGKELELLAILPDNNGSL 637
 QY 482 YDAIKKYLCTDCPTSPQCVAR---TLGQQTVAALATKIALQNNCKMGE-----L 530
 DB 638 YGDLKRICETELGLISQCLTKHVFKISKQ---YLADVSLKINVMKGRNVTYLDNIS 692
 QY 531 WRV---DIPKLVTAVIGDCHMTA--GRSITAGFASIN-BGMTRMPCRCIFDRGQ 583
 DB 693 CRIPVSDIP--TIFGADVTHPENGESSPSIAAVALSQDMEVTKIAGLVCAQAHKQ 749
 QY 584 ELVLDLKYCLQALALR-----ANNSCHEWMSRIIVYRDVGQDLTKTLVNYE 630
 DB 750 ELIQDLYKTMQDPVNGYVSGMIRDLISFRKAGQKPLRIIFRDVSGEQFQVLYLE 809
 QY 631 VPQFLDCKSIGRGYNPRLVIYVKKRVNTRFPA-----OSGRLQNPFGYIYDEV 684
 DB 810 LDAIRKACASLEPNQPPVTFIVQKRHTRLFPANNHRDNKSTRSGNIIPLGYVADTKIC 869
 QY 665 RPEWDEFTISQAVRSGSVPTHYVVIYDNGSLKPDHIGLTYLCKHYVMPGVIRVPA 744
 DB 870 HPTEDFTLCSHAGIGTISRPAHIVLMDENNFRADGIGSLTNMLCYTAACRSTSVELIP 929
 QY 745 PCQYAHKLAFLVQGISHREPNLISLN 770
 DB 930 PAVYAHLAFA--RFYLEPEIMQDN 953

RESULT 9
 A:Agonate (AG01)-like protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: A84668
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unagun, L.;
 Neus, D.; Nielsen, W.C.; White, C.; Eise, J.A.; Salzb, S.L.; Frazer, C.M.; Venter,
 Nature 402, 761-768, 1999
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84668
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-930 <STO-
 A:Cross-References: GB:AE002093; NID:9388534; PIDN:AACT7862.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: Atg27040
 A:Map position: 2
 C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 11.9%; Score 488.5; DB 2; Length 930;
 Best Local Similarity 21.3%; Pired. No. 3.8e-27;

Matches 185; Conservative 170; Mismatches 340; Indels 173; Gaps 34;
 QY 20 GSGGIVLSTNHFEL-TSRPOMALYQYH-----DYNPLWEARLRALLFQ-----HE 68
 DB 63 GTRGQKIPLTTHFHVADVANLQGHFHHYSVALFYDDGP-VEQKVGKILDKHQTYS 121
 QY 69 DLIGKHAPDG-----TIFLPRQLQ--QKTEVFSKTR-----NG----- 102
 DB 122 DLDGEFAIDGSKILFTYGAIDPSNMDDSVYLEBASISKQFVRANGSGPNQESPSD 181
 QY 103 -----EDVRIITLTNEIP-----PISPTCLQYNIIFRLKIM 137
 DB 182 GDRKLRRENRKSNRVEISYAKIPLQALANAMQSESENSQAIRVDIILRQHAARQ 241
 QY 138 NLQIGRNYNPNNDPIDP--SHRLVWPGFTTSIIQYENSIMLCTDV-SHKVRSBETVL 194
 DB 242 GCLIVRQSPFH-NDPINCPEVGNILGCRGHSSFRITQGSMLMDVTTMIIKPFVY 300
 QY 195 DMENFYTQTEEHKQ-EOVSKELIGLVLTAKNKTAVDDIDMDQPKSTFK----- 247
 DB 301 DFLIANQNAIDPYSIDWSKAKRTLNKRVKSPSQGEFTIGLSDKPREQTEFLKKNP 360
 QY 248 ----KADGSEVPLEYRK-----QYNOETDLKQPVLSQPKRRRGPGTLPGRAMLI 297
 DB 361 NENGEFETTEVADVDFRTRHIDQYSADLCIN---VGKPKR-----PTIIP 406
 QY 298 PELC-----YLTGLTRRNDFNVMKDLAVHRLTPEORREVRGLIDYTHQDNVOR 350
 DB 407 LELCALVPLQRYTKALTFRQS-----ALVEKRSQKQERMTVLSKALKV--SNYDAP 458
 QY 351 ELRDMGLSPDNLSSGRILOTETKHQ--GGKTDVNOQFADWSKETRGADLISVPEPD 408
 DB 459 LRSQGISISSNFTQVEGVRLPAPKMGCSSEITPRNKRMPNKE-----FVEPTKIQ 513
 QY 409 NMLIYTRNRYA--ANSLIONLFKVTAPAMQMRKALMIEVD-----RTEAY 455
 DB 514 RMVYV---NFSARCNVRQVVDLTKIGSKGIEIASPQVFEENQOFRRAPMIREVM 569
 QY 456 LRVLQKVTADQIVVCLLSNR-KDKYPAIKYLCDCPTSPQCVVATLKKQGVNMI 514
 DB 570 FKDIQGLKPGVQFICVLPDKKNSDLYGPMKKULTTEGIVTQCAPTRQPDQL--- 626
 QY 515 ATKIALQNNCKMKG--ELMRVD-----IPLKLVITVIGDCHMTAGRRSIAGV 562
 DB 627 -TNLLIKINAKGLNSMLSVETRAFYISNVP---TILLMDVSHG-SFGSDVPSIA 681
 QY 563 ASINEG---MTRWESRCIFDQSGELVDGL-----KYCLQALPAMNSCHEWMS 609
 DB 682 AVVSRREWPLISKYRASVTRQPSKAMIESLVKRNKGTEDDGIKELVDVYSSNRKPE 741
 QY 610 RIIVYRDVGQDLKTLVYVEVPOFLDCKSIGRGYNPRLVIYVKKRVNTRFPAQSGR 669
 DB 742 HILFRDVGSGSFQVNLVIELDQIIEACKLIDANNPFLVIAQKNHHTFQPTSP- 800
 QY 670 LQNPFGYIVDEVTRPEWDEFTISQAVRSGSVPTHYVVIYDNGSLKPDHIGLTYL 729
 DB 801 -ENVPFGTIIIDNKICHPKNDPFLCAHAGMIGTRPTHVHYVDEIGFADELQELVHSL 859
 QY 730 CHYVMPGVIRVPAQCYAHKLAFLVQGISHREPNLISLN 770
 DB 860 SYVQSTSAISVAVICTAHLAAGQG 887

RESULT 10
 JC6569
 Translation initiation factor eIF-2C - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
 C:Accession: PC6505; JC6569
 R:Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
 Gene 211, 187-194, 1998
 A>Title: Molecular cloning and characterization of a rabbit eIF2C protein.
 A:Reference number: JC6569; MUID:98267198; PMID:9602122

A/Accession: PC6505
 A/Molecule type: protein
 A/Residues: 336-359/380-694-711 <ZOU1>
 A/Experimental source: liver
 A/Accession: J06569
 A/Molecule type: mRNA
 A/Residues: 1-813 <ZOU2>
 A/Cross-references: GB:AF005355; NID:g253158; PIDN:AC2423.1; PID:g3253159
 C/Species: rabbit translation initiation factor eIF-2C
 C/Suprafamily: rabbit translation initiation factor eIF-2C
 C/Keywords: liver

Query Match 11.8%; Score 482; DB 2; Length 813;
 Best Local Similarity 23.9%; Pred. No. 9.1e-27;
 Matches 191; Conservative 147; Mismatches 341; Indels 120; Gaps 32;

39 PGMALYQYHIDNPMFMEARRLSAL---LFGH---EDLIG-KCHAPDG-----TILFPR 87
 4 PKIDYHIELDKPEKCRVRVREIVEHMGHFKQIGDKRPVDSGRKNLYTMPLPIG 63
 88 LQKQYEVFSKTRNGED--VRIIT-----TINELPPTSCQFYNIIFRR 132
 64 REKYLEVELTLEGEKDIIFKYSIKWBSVSLQALHDLSGRLPSVPETIOLADVMWRH 122
 123 LKINNLQOIGRNNYNNPNDIPIDP-SHRLVIMPFTTSILOEYNSIMLCTDVSFKVL-RS 190
 123 -LPSNRVYFPGRSFPTASGCSNPLGGGREGVFGHQSVRPSLMKMLNIDVSATAPYKA 181
 191 ETVLDPF--MNFYHQTENK-----FOEYVSKELIGLVITKY---NNKTYRVDIDMD 239
 192 QVIVFVCEVLDLFEKISIEQKPLTDSQRYKFKIKGKVEITHGQMKRKRYCNVTRR 241
 240 QNPKEFPKADS---EVSFLEYRKYQNGEITLQKPV--VSPKRRRPGGLTPBA 294
 242 PASHQTFPLQSGSGQVTECTVAQYFRDKRKLVLRYPHLPCQLQVGEQNH-----T 291
 295 MLIPELCYLTG-----LTDKRNDFVMKDLAVHTLTPQORREYGRLLIDYHKN 347
 292 YLPFLVCNIVAGQRCIKLTD---NQTSM--IRATASAPD-ROEISKLMR--SASFN 343
 348 VORELDWGLSPDSNLSFSGRILOTETKHOGGTFDYNPQPADMSKSTRAPLISVPL 407
 344 TDPYRREFGIMVKDEMTDVTGRVLQPPSILYGR-----NKALATPVGVDWM 391
 408 DN-----WLLIYTRANYEA---NSLIONLFKVTPEAMGMOKR---AIMIEV 448
 392 RKKQHTGIEIKVMALACAPORCTEVHLSFTQOLKISPDAGMPICQGPCFKVAG 451
 449 DRTAYLRVLOQKVTADTQIVVCLLSNRKDKYDAIKKYLCTDCTPSPQCVANTLQK 508
 452 ADSVGMFRHL-KNTYAGLQLVVIL--PKTPTVYAEVKEVGDVTLGMAATQCCOMKNV--Q 507
 509 QTVMAIATKIALQNNCKKGG-----ELMRVDIPKLWIVISIDCYHMTG---RSIA 559
 508 RTTPQTLNLCIKINVKGGVNNIILPGGRPEVPOQVIFLQADVTHP-PADGKKP 566
 560 GYVASINEMTWFSRCIFQDRGOELVLDGLKVCLOALRAMNSCNEVPSRIIYVRDVG 619
 567 AYVGSNDAPNNYCATVRVQQRHQLIDOLAMVRELLIQFYKSTRPRTIRIFRDVVS 626
 620 DQGLKLVNVEVPQPLDCKSIGRGVNPILYIVYKKNVRFPA-----QSGRLQNL 675
 627 EOPQOVLNHEHLAIERACIKLEKYQDPGITIVQKRHTRLFTDKNERYVSGKNLPA 686
 676 GTVIDEATREPMYDEFFIVSOAVRSGSVPTHYVYVYNSGLKPHIQLTYKLCHIIYN 735
 687 GTTVDTIKIHPTREFDYLCSHAGIQCTSRPSHYVLMMDNRSSDELQILTYQLCHTYR 746
 736 WEGVTRVVPAPQYAKLAF 754
 747 CTRSVSIPAPAYAHVAF 765

RESULT 11

T01113
 Translation initiation factor eIF-2C homolog T21114.12 - Arabidopsis thaliana
 N/Alternate names: Argonaute (AGO1)-like protein
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Dates: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
 C/Accession: T01113; D84739
 R/Romney, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 submitted to the EMBL Data Library, December 1997
 A/Description: Arabidopsis thaliana chromosome II BAC T21114 genomic sequence.
 A/Reference number: Z14209
 A/Accession: T01113
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-887 <ZOU>
 A/Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702284
 A/Experimental source: cultivated Columbia
 R/Lin, X.; Kaul, S.; Romney, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Talion, L.;
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: AB4420; M0ID:20083487; PMID:1061197
 A/Accession: D84739
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-887 <STO>
 A/Cross-references: GB:AE002093; NID:g2702284; PIDN:AB91987.1; GSPDB:GN00139
 A/Genetics:
 A/Gene: At2g32940; T21114.12
 A/Map position: 2
 A/Intons: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3; 5
 C/Suprafamily: rabbit translation initiation factor eIF-2C

Query Match 10.6%; Score 434; DB 2; Length 887;
 Best Local Similarity 22.2%; Pred. No. 3.2e-23;
 Matches 198; Conservative 153; Mismatches 349; Indels 190; Gaps 41;

3 FGVTNRQNLDHVESKTSKSGIIVRLSTNHFRTSR-POMALYQYHIDNPMFMEARR 57
 24 YDITTRRGV-----GTTGNPIELCTNHNFNVSVPQDVVFQYVSYITTEGDAVDGT 75
 58 RLRSALIFQ-----HEDLICKAPFG-TLFLPKLQK-----YREVSXTNG-- 102
 76 GISRKMDQLEFKTYSDDLQKRLAYDGEKTLVYVGPLPQNEFDPVIVYSGFSKDDGVS 135
 103 -----EDVRITITLNLPL-----PISPTCLQFYNI 129
 136 DGGSSGTCRSKRSRFLPRYKQIHAAEIPKTVLGTQRGAVTPDKAQDALRVLDIV 195
 130 FRRLIKMINLOQIRN-----YNNPNDIPIDISHRLVYWPQTTSLQYENSIMC 180
 196 LR-----QAAERGLLVROAFPHSDGHPKVGAGVGI-RGLHSRPTFGGLSLN 246
 181 TDVS-HKYLESEVLDPMFNFYHQTENKQ-----EVSKEILGLVLTLYKKNKYRVD 235
 247 IDVSTWILEPGLVTEFL-KANQSVETPRQIDMIKVAAMLGAMVYKATHRMPEPT-- 302
 236 IDWDNPKS-----TFKKADS-----EVSFLEYRKYQNGEITLQKPV--VSPKRR 284
 303 IGLSKRCNOQLPSMKIKDGEREVPRIETIVDYFQOTYEPISASVFFCLDVGKDR-- 360
 285 GPGGTLPGPMLIPELCYLTGLDKNRDNWKKDLAVNHR--LTPEQORREYGR----- 337
 361 -----PNTLPLEFNLSLO-----RYKPLSGRQVULVSSRKPRLKIKTLN 405
 338 -LIDYHKNNDYQRELWGLSPDSNLSFSGRILOTETKHOGGTFDYNPQPADMSKE 395
 406 DAHTTYCYDQDPF--LAGGGISIEKEMTQVEGRVLKPPML-KFGKXNEDFQPCNGRNFN 461
 396 TRGAPLISVPLDNWMLIYTRYAYE--ANSLIONLFKVTPEAMGMOM-RKALMIEVD-- 449
 462 NK-MLEPRAIKSMALV--NFSPPCDSSHISRELLISGMRKGLFIDRPPALVEEDPQ 515

QY 450 -----DREAYLRVLOQKVTADTOIIVCLISSNR-KDKYDAIKKYLCTDCEPTSCQCV 501
 Db 516 YKXAGPEREKXIAMTKLFPDPHILICILPERKTSIDYGPWKICILTEGIIHQIC 575
 QY 502 ARLVKGQVMAIATKALQNMCKG-----GELMRVIDPL-KL-VNIVGDCVDMFT 552
 Db 576 PIKISDQ-----YLVNVLKINKSLGINSLEIEFSYINPLINKPTLILGMDVSHG-P 629
 QY 553 AGRRSIAQVAFASINEG-----MTWFSRCPQDRGOELVLDGLKVCLOALRAWNS-CNEYM 607
 Db 630 PGADVSAVAAYVGSCKMPLISRYRAVRQSPRLMIDSLFQPIENTKGDGMKINELF 689
 QY 608 -----PSRIYRDVGVDGQKTLVNTVEVPOFLDCKLSIGRGVNPRLTYIVVVK 656
 Db 690 VEFYRTSRARKPKOIIIFRDGVSESOPEOVLKIEVQOIKAYORLESVDVPEKFTVVAQK 749
 QY 657 RVNRFPAQSGGRQONPLPGTVLDVEVTRPEWYDFIVSOAVRSGSVPTHYVIVYDNG 716
 Db 750 NHHYKLVQAKGP-ENVPAQIVVDTKIVHTNYDFTVCAPAGKIGTSRPAHNVLDDEIG 807
 QY 717 LKPDHIOF-----LTYKLCIHYVWMPGVIRV-----PABQYAHKLAFLVQ 758
 Db 808 FSPDQLNLHLSLSYKILNSIFVNSILLCVFVLSVAPVRVYAHLLAAQVQ 857

RESULT 12

T23164
 hypothetical protein T22B3.2a - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C/Accession: T23164; T25099
 R/Cottage, A.
 submitted to the EMBL Data Library, January 1996
 A/Reference number: Z19701
 A/Accession: T23164
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1032 <MTL>
 A/Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a
 A/Experimental source: clone K01A6
 R/Lennard, N.
 submitted to the EMBL Data Library, December 1995
 A/Reference number: Z19981
 A/Accession: T25099
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1032 <MTL>
 A/Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a
 A/Experimental source: clone T22B3
 C/Genetics:
 A/Gene: CESP:T22B3.2a
 A/Map position: 4
 A/Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1
 C/Subfamily: rabbit transfection initiation factor eIF-2C

Query Match 10.5%; Score 429.5; DB 2; Length 1032;

Best Local Similarity 22.1%; Pred. No. 8.6e-23;
 Matches 203; Conservative 144; Mismatches 367; Indels 203; Gaps 36;

QY 20 GSSGIVRLSTNHFPL-TSRPOMALYOHID-VNPLMABRL-RSALLP-----QHED 69
 Db 88 GTIGKIPKSNFPAVDLKNPKNVAVQYHVEVHH-DGCKLDNDENHIIWKAVSDHPN 145
 QY 70 LIGK--CHAFDGT-LILFPLKRLQ-----QKTVESKTRNGEDVRIT----- 109
 Db 146 IHHNKALAYDGHQYVARLEFPDQGSVRLDEALTPKONRDRCAISQNVGFLV 205
 QY 110 -----TLNLELPPTSTGCFYNIIFRLIKIMNQQIGRNVN-----PNDPI 153
 Db 206 LEMQRRITNMDERVLTPRIQIILICROSLTCLPLKN-SANVTYKSSCYRPTAQAAL 264
 QY 154 DIPSHLVIMPGETTSIIQYEN-SIMLCTDVSHKYLRSB-TVLDPMNFYHQ----- 203

Db 265 DLBGGK-EMMTGTFSSAHISNVRPLNIDVANHAFYKTRITVLQFWCDVILNERTSKPNR 323
 QY 204 -----TEH 207
 Db 324 NNRPQFAGYRGARGGARSGSYONFANRPPGAVNRDDFGNGLFTYDTLSDRQOLS 383
 QY 208 KFGQVSKELI-GVYLTKNKNTYRYVDIDMDQNP--KSTFKKAD--GSEV--SFLERY 260
 Db 384 SFETTFGDSIRGKIRATHRPNARIVYKNSLQPLADKLMFGQIDBEGQVGVSDYF 443
 QY 261 RKOYNQBITDKQPVLYSOPRRRGPGTLPGPMLIP-ELC-----YLTGLTDKRN 312
 Db 444 SEKG-----FLKYPRL-----PCLHVGSP-----PTRNIFLPMBCLLIDSPOKRNKMTKQTS 492
 QY 313 DPNVKKDLAVNTRLPBROREVRLLDYIHKNNVORRELDWGLSPDSNLSFGRILO 372
 Db 493 --AIIKAAVA--TQREDRIKOLA--QAASGTDBFLKEFGVAVSSQMIETSAVIO 544
 QY 373 TEKIHGGKTPDYN-----POFADWSKETRGAPLISVXPDLNMLIYTRRYEAANSLION 428
 Db 545 PPIIFEGGNNSINPVFPKDGSMMDHQLYMPATCRSYEMIALVDPDQTSIQTFQCS 604
 QY 429 LFKVTPANGQMRK-----ALMIEVDRTAVLRVLQOKVTADT-QIYVC 472
 Db 605 LTMKATAMGANFPWPDLVYKGRSKEDVCTLFETIAD-----EYRVNTYQDCIIV 655
 QY 473 LLSNRKQYDAIKKYLCTDCEPTSCQCVAVARTLGKQVTMAIATKALQNMCKMG----- 528
 Db 656 VIGSNISDIYMTVWQSDIYVIGINSQCYLKNKVR--PTPANCANIILIKNMKMGINSR 713
 QY 529 -----ELMRVIDPLKVMIVGIDCYDMNAGR-----STAGFVASINEGTRWFSRCI 577
 Db 714 IVADQITNKYLVDQ-----TMVVGIDVTPQOEMKMNPSVAALVANDLBPQSGAVK 770
 QY 578 FQDRGOELVLDGLKVCLOALRAWNSCNEYMSRIIVRDGVGQGLKTLVNTVEVPOFLDC 637
 Db 771 VQKCRRESVYLLDIRRILIFFYHHTQKPARILIVYDVGSEQFSVLEELIQSIRTA 830
 QY 638 LKISRGVNPRLTYIVVKKRVNTRFPAQ-----SGRLQNPPLPGTVLDVEVTRPEWYDFI 693
 Db 831 CLAIADFPFPITVYVQKRRHAR-FCFEPNDVWAKRANVPGLTVDGIYSBGFDFYL 890
 QY 694 VQGNVRSQSVSTHNVIVDNGSLKPDHIORLTYKLCIHYVWMPGVIRVPAQYAHKLA 753
 Db 891 CSHYVQGTSRPARHNVLDDECKFTADEIQNTTYGMCHYGRCTSVSIFTVYADIVA 950
 QY 754 FIVGGSIHREPNIUSN 770
 Db 951 TPARCHIKRKLGLADNN 967

RESULT 13

T23165
 hypothetical protein T22B3.2b - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C/Accession: T23165; T25100
 R/Cottage, A.
 submitted to the EMBL Data Library, January 1996
 A/Reference number: Z19701
 A/Accession: T23165
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1035 <MTL>
 A/Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b
 A/Experimental source: clone K01A6
 R/Lennard, N.
 submitted to the EMBL Data Library, December 1995
 A/Reference number: Z19981
 A/Accession: T25100
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1035 <MTL>

A:Cross-references: EMBL:Z68300; PTDN:CAA92619.1; GSPDB:GNO0022; CESP:T22B3.2b
 A:Experimental source: clone T22B3
 C:Genetics:
 A:Gene: CESP:T22B3.2b
 A:Map position: 4
 A:Inserts: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 796/2; 964/2; 1006/1
 C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.5%; Score 429.5; DB 2; Length 1035;

Best Local Similarity 22.1%; Pred. No. 8.6e-23;

Matches 203; Conservative 144; Mismatches 367; Indels 203; Gaps 36;

QY 20 GSSGIIVLSTNHFRL-TSRPQWALYOHID-YNPLMEARL--RSALIF-----QHED 69
 DB 91 GTTGKRIPVKSNFADLKNPKVAVVQYHVEIHP--GCRKLKDDEKRIIFKAVSDHPN 148
 QY 70 LIKG--CHAFDGT-ILFLPKRLQ-----QKTVESKTRNGEDVRIIT----- 109
 DB 149 IFHNKFAIADGAHQLYTVALEFPDDQSVRLDCEATLPKXNRDRTRCAISQNVGVL 208
 QY 110 -----TLTNELPPTSPCLQFNYIIFRLIKIMLQOIGRNYN-----PNDPI 153
 DB 209 LEMQRTTNLDERVLPTQIIDLICRQSLTCELLKN-SANFTYKSSCYRIPTAAGAL 267
 QY 154 DIPSHRLVWPFGFTTSILOYEN-SIMLCTDVSHTKLRS--TVLDPMFNHYO----- 203
 DB 268 DLEGGK-EMWTGFFSSAHIASNYRPLINDVAHTAFYKTRITVLOQFMCVLERISKPNR 326
 QY 204 -----TDEH 207
 DB 327 NNPRGPGAPGGRGGRGAGGSYONFGNRPGRANVDDFGNGLFTMDTSLRDTLS 386
 QY 208 KEOQVSKELI-GLVLTLYNNKTYRVDIDWDQNP--KSTFKKAD--GSEV--SFL 260
 DB 387 SEETLIFGDSIGMKIRATHRNPAIRVYKNSQLPADKLMFGQIDEGRQVVCVADYF 446
 QY 261 RKQNOEITDLKQVLYSQPKRRRPGGTLPGRAMLIP-ELC-----YLTGLTDKRN 312
 DB 447 SEKYG--PLKYPKL--PCLHVP--PTRNIFLPMHCLIDSPQKYNKMEKQTS 495
 QY 313 DENWMDLAVHTRLPQORQREVGRLIDYHKNDNVQRELADWGLSPDNLSPSGRLQ 372
 DB 496 --AIIKAAVDA--TQREDRIKOLA--AQAISGTDPLKEFGVAVSSQMIETSAVIO 547
 QY 373 TEKIHGGKTFDYN---PQFADWSKETRGAPLISVPELDNMLIYTRRNEAANSLION 428
 DB 548 PPPIMGSGNRRSINPVVFPKXGSMWMDHOTLYMPATCRSSYMIALVDPDDQTSIQTFQS 607
 QY 429 LKRVTPAMQWQK-----AIMEVDRTREAYLRVLOQKVTADT-QIYVC 472
 DB 608 LTKATATAMGNFPRMVDLVKYGSKEDVCTLFTEIAD-----EYRVNTVDCDCLIV 658
 QY 473 LLSNRKDKYDAIKKYLCTDPTSPQCVARTLKQOVTWATATKIALQMNCKMG-- 528
 DB 659 VLOSKNSDLYMTVYKESQDIYHGMSQVLMKNVSR--PTPATCANIILKMKMGKINSR 716
 QY 529 -----ELMRVDIPKLVMIVGIDCYADMTAGRR--SIAGFVASINEGTRMFSR 577
 DB 717 IYADQITNKYLVDP--TWVVGIDVTHPTQAEKRMAMPVSAIYANVDLLPQSYGANV 773
 QY 578 PQORGEIVDGLKVCQALMAANSQENYMPRSIIYRDVGSGQKTLVNEVPOQLDC 637
 DB 774 VQKCKESVYTLDAIRERITTYRTKQKPAIIYVRDVSSEGSQSVTLREIQSIRK 833
 QY 638 LKISIGGYNRLTVIVKQKAVNTRFPAQ--SGGLQNPPLPCTVIDEVTRPEWYDPT 693
 DB 834 CLAIADFPRLPTIYIVQKSHARIFCKFENDWVGAKVPPQTVDTGIVBEGDFYL 893
 QY 694 VSGQVNSGSVPTHYVIVYNSGLKPDHQRLLYKLCITYYMPVGIYRPAPOQYAHKA 753
 DB 894 CSHYGVQGSRRPARVHLLDECKFTADEIONITYGCHYVGRCTRSVSIPTVYIADLYA 953
 QY 754 FLVGSIIHREPNTLSN 770

DB 954 TRACHIKKTLGLADNN 970

RESULT 14

S41013
 hypothetical protein ZK757.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000

C:Accession: S41013

R:Thomas, K.
 submitted to the EMBL Data Library, December 1993

A:Reference number: S41011

A:Accession: S41013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-958 <THO>

A:Cross-references: EMBL:229121

C:Genetics:

A:Inserts: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; Length 958;

Best Local Similarity 21.6%; Pred. No. 3.2e-22;

Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;

QY 20 GSSGIIVLSTNHFRL-TSRPQWALYOHID-YNPLMEARL--RSALIF-----QHED 69
 DB 11 GTTGRIPVKSNFADLKNPKVAVVQYHVEIHP--GCRKLKDDEKRIIFKAVSDHPN 68
 QY 70 LIKG--CHAFDGT-ILFLPKRLQ-----QKTVESKTRNGEDVRIIT----- 109
 DB 69 IFHNKFAIADGAHQLYTVALEFPDDQSVRLDCEATLPKXNRDRTRCAISQNVGVL 128
 QY 110 -----TLTNELPPTSPCLQFNYIIFRLIKIMLQOIGRNYN-----PNDPI 153
 DB 129 LEMQRTTNLDERVLPTQIIDLICRQSLTCELLKN-SANFTYKSSCYRIPTAAGAL 187
 QY 154 DIPSHRLVWPFGFTTSILOYEN-SIMLCTDVSHTKLRS--TVLDPMFNHYO----- 203
 DB 188 DLEGGK-EMWTGFFSSAHIASNYRPLINDVAHTAFYKTRITVLOQFMCVLERISKPNR 246
 QY 204 -----T 204
 DB 247 NNPRGPGAPGGRGGRGAGGSYONFGNRPGRANVDDFGNGLFTMDTSLRDT 306
 QY 205 EHNFOQVSKELI-GLVLTLYNNKTYRVDIDWDQNP--KSTFKKAD--GSEV--SFL 257
 DB 307 QLSSEFTRIFGDAIRGKIRAPRAIRVYKNSQLPADKLMFGQIDEGRQVVCVADYF 366
 QY 258 EYRKQNOEITDLKQVLYSQPKRRRPGGTLPGRAMLIP-ELC-----YLTGLTDK 309
 DB 367 DYSEKKG--PLKYPKL--PCLHVP--PTRNIFLPMHCLIDSPQKYNKMEKQTS 415
 QY 310 MRDNFMKDLAVHTRLPQORQREVGRLIDYHKNDNVQRELADWGLSPDNLSPSGR 369
 DB 416 QTS--AIIKAAVDA--TQREDRIKOLA--AQAISGTDPLKEFGVAVSSQMIQTTAR 467
 QY 370 ILQTEKHGGKTFDYN---PQFADWSKETRGAPLISVPELDNMLIYTRRNEAANSL 425
 DB 468 VIQPPPIPFQGNNSVAVPVVFPKXGSMWMDHOTLYMPATCRSSYMIALVDPDDQTSIQTF 527
 QY 426 IONLFKTPAMQWQK-----AIMEVDRTREAYLRVLOQKVTADT-QI 469
 DB 528 QOSLTKATAMGNFPRMVDLVKYGSKEDVCTLFTEIAD-----EYRVNTVDCDCLIV 578
 QY 470 VCLLSNRKDKYDAIKKYLCTDPTSPQCVARTLKQOVTWATATKIALQMNCKMG-- 528
 DB 579 IIVLOSKNSDLYMTVYKESQDIYHGMSQVLMKNVSR--PTPATCANIILKMKMGKINSR 636
 QY 529 -----ELMRVDIPKLVMIVGIDCYADMTAGRR--SIAGFVASINEGTRMFS 574
 DB 637 NSRIVADKITNKYLVDP--TWVVGIDVTHPTQAEKRMAMPVSAIYANVDLLPQSYGA 693

QY 575 RCIFDROGELVDGLKVLQALRAMNSCNEWMSRIIVYRDVGQDLKTLVNYEVPQF 634
 DB 694 NVKVKCKRESVYLLDAIRERITTFYRHTKQKPARIIYRDVSGQSEVLRRELSQI 753
 QY 635 LDCLKSIGRGNPRLTVIVKRVNTRFFAQ---SGRLQNPPLPGTVIDVEVTRPEWYD 690
 DB 754 RTACLAIAEDFRPPIITYIVQGRHARIFCKYQNDWVGAKAVPFGTTVDGIVSEGF 813
 QY 691 PFIVSQAVSGSVPTHTNVIYDNGSLKPRHIGLTYKLGHIYNNMGVIRVAPCOYAH 750
 DB 814 FYLSHVGQTSRPARHVLDECKFTADEISITIGMCHYGRCTRSVSIPTVYIAD 873
 QY 751 KLAFLVGSIHREPNLSLN 770
 DB 874 LVATRARCHVKKLGLADNN 893

RESULT 15
 DB8568
 protein ZK757.3 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: DB8568
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_eleg
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: DB8568
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1040 <STO>
 A:Cross-references: GB:chr_III; PIDN:CA82941.1; PID:93877004; GSPDB:GN00021; CSDP:ZK757
 C:Genetics:
 A:Gene: ZK757.3
 A:Map position: 3
 C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; length 1040;
 Best Local Similarity 21.6%; Pred. No. 3.6e-22;
 Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;

QY 20 GSSGIIVLSTNTHFL--TSRQMALYOHID--VNPIMEARL---RSALF-----QHED 69
 DB 93 GTIGRIQIVKSNFPMADIKNPKCVVIOYVEIHHP--GCRKLDXDEKRITFWKAVSDHPN 150
 QY 70 LIGK--CHAFDGT--ILFLPKRLQ-----QKTEVESKTRNGEDVARIIT----- 109
 DB 151 IFHNKFALAYOGAHLVYVARLEFPDQGSVRLDCEASLPRKNDRTCAISIQNGFVL 210
 QY 110 -----TLNNEZLPPTPCLOFYNIIFFRLKIMNLQOIGRNVN-----PNDPI 153
 DB 211 LEMORTNNLDERVLTPIQIIDIICRQSLTCPLKN--SANPYTWKSSCYRIPTAAGAL 269
 QY 154 DIPSHRLVYWPGETTSILOYEN--SIMLCTDVSHKVLASE--TVLDFMENFYHQ----- 203
 DB 270 DIEGK--EMWIGFSSAHASNYRPLNIDVAHTAFYKTRITVLQFMCDVNERTSKPNR 328
 QY 204 -----T-----T 204
 DB 329 NNPRGPGGPGGPGYRGGRGGSGYGNFNGRPGGANVRDFFGNGLFTMDTLSDRT 388
 QY 205 EHHKFOEQYSKELI--GLVVLTKYNNKTYRVDIDMDONP--KSTFKAD--GSEY--SFL 257
 DB 389 QUSSEETRIFGALIKMKTRAHRRPNAIRVYVNSLOLPADKLMFGQIDBGRQVYCSVA 448
 QY 258 EYRKQYNOEITDLKQPVIVSQPKRRRGSGTLLPGPAMLIP--ELC-----YLTGLTDK 309
 DB 449 DYFSEKYG---PLKYPKL---FCLHVG---PTRNIFLPMEHCLIDSPQKYNKQMSK 497
 QY 310 MENDFVMDLAVHTRLTPEGRQREGRRLDIYHNKDNVQRELBDWGLSPDSNLLSPSGR 369

DB 498 QTS--AIKAAVDA-----TQREDRIKOLA--AQASFQIDPEFLKFGVAVSSOMIQTRAR 549
 QY 370 ILQTEKINGGKTPRYN-----POFADWSEKRETPALISVKPLDMNLLITRRNYEAANSI 425
 DB 550 VIQPPPIFGANNRNVNVPFPGKGSWTMDQUTLMPATCRSYSMLALVDRDQTSLOTF 609
 QY 426 IONLFKVTYPMQOMWK-----AIMIEVDRTYAVLYLQOKRTADT-QI 469
 DB 610 CQSLTMKATAMGNPFPEWBDLVKYSKEDVCTLFTETIAD-----EYRVNTEVDC 660
 QY 470 VYCLLSNRKDYDAIKKYLCIDCEPTPGQCVVARTIGRQVMAATATKALQMNCKXGG- 528
 DB 661 ITIVLQSKNSDIYMTVKQSDIVAGIMQCVLAKVSR--PTPATCAITVLKLNKMGGI 718
 QY 529 -----ELMRVDIPLKLMIVIGIDCYHDMTAGRR---SIAGFYASINEGTRWFS 574
 DB 719 NSRIVADKTIKKYLVDP---TMVVGIDVTHPTQAEKMMNPSVAIAVANDLLPQSYGA 775
 QY 575 RCIFDROGELVDGLKVLQALRAMNSCNEWMSRIIVYRDVGQDLKTLVNYEVPQF 634
 DB 776 NVKVKCKRESVYLLDAIRERITTFYRHTKQKPARIIYRDVSGQSEVLRRELSQI 835
 QY 635 LDCLKSIGRGNPRLTVIVKRVNTRFFAQ---SGRLQNPPLPGTVIDVEVTRPEWYD 690
 DB 836 RTACLAIAEDFRPPIITYIVQGRHARIFCKYQNDWVGAKAVPFGTTVDGIVSEGF 895
 QY 691 PFIVSQAVSGSVPTHTNVIYDNGSLKPRHIGLTYKLGHIYNNMGVIRVAPCOYAH 750
 DB 896 FYLSHVGQTSRPARHVLDECKFTADEISITIGMCHYGRCTRSVSIPTVYIAD 955
 QY 751 KLAFLVGSIHREPNLSLN 770
 DB 956 LVATRARCHVKKLGLADNN 975

Search completed: May 5, 2004, 15:14:48
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:10:00 ; Search time 52 Seconds
(without alignments)
4702.435 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102
Sequence: 1 MIFGVNTRQNLDHVKESTG.....VGQSIHREPNTLSNRLYYL 775

Scoring table: BLOSUM62

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
```

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Database :

```

1:  sp. arcea.*
2:  sp. bacteria.*
3:  sp. fungi.*
4:  sp. human.*
5:  sp. invertibrate.*
6:  sp. mammal.*
7:  sp. mhc.*
8:  sp. organelle.*
9:  sp. phase.*
10: sp. plant.*
11: sp. rodent.*
12: sp. virus.*
13: sp. vertebrate.*
14: sp. unclassified.*
15: sp. virus.*
16: sp. bacteriap.*
17: sp. archaeap.*

```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4102	100.0	775	4	O96J45	O96J45 homo sapien
2	4087	99.6	861	4	O96J94	O96J94 homo sapien
3	4072	99.5	861	4	O8BYT5	O8BYT5 homo sapien
4	4071	99.2	861	4	O95J04	O95J04 homo sapien
5	3980	97.0	862	11	O9UM37	O9UM37 mus musculus
6	3883	94.7	829	4	O8NA60	O8NA60 homo sapien
7	2828.5	69.0	858	13	O8UVX0	O8UVX0 brachydanio
8	2144	52.3	882	4	O7ZK23	O7ZK23 homo sapien
9	2103.5	51.3	852	4	O8NH32	O8NH32 homo sapien
10	2101.5	51.2	852	4	O8NV88	O8NV88 homo sapien
11	2098.5	49.0	852	4	O7KZ34	O7KZ34 homo sapien
12	2009.5	45.0	854	5	O9GDA8	O9GDA8 strongyloce
13	1859.5	45.3	878	11	O8GCT6	O8GCT6 mus musculus
14	1838.5	44.8	666	4	O8N6G9	O8N6G9 homo sapien
15	1811.5	44.2	808	5	O9GPA7	O9GPA7 strongyloce
16	1766	43.1	971	11	O9UM66	O9UM66 mus musculus

17	1764	42.3	0	971	11	O8CDG1
16	1754	42.8		973	4	O8TC59
15	1468	35.8		866	5	O76B22
20	1389.5	33.9		580	11	O9WMV6
21	1295.5	31.6		824	5	P90T86
22	1230	30.0		530	4	O9ESM6
23	1209	29.5		421	5	O9GPAS
24	1162	28.3		722	5	O17567
25	1059	25.8		979	5	O8MMJ6
26	1009	24.6		1208	5	O8EHT5
27	989	23.9		791	5	O8ISG8
28	981	23.9		791	5	O8ECV7
29	962	23.5		371	4	O9NMZ8
30	958.5	23.4		373	13	O7ZMB5
31	804.5	19.6		780	5	O8MKZ9
32	803.5	19.6		780	5	O8MQC1
33	728	17.7		781	5	O9UJC9
34	537.5	13.0		1058	10	O8SIR2
35	520.5	12.7		1194	10	O9SHP2
36	508.5	12.4		860	11	O8CTG0
37	508.5	12.4		1014	10	O9SHF3
38	507	12.4		978	10	O8LPU0
39	505.5	12.3		891	5	O8EBJ3
40	505.5	12.3		910	5	O16720
41	504.5	12.3		861	11	O8CPJ8
42	503	12.3		896	10	O8AYI4
43	502.5	12.3		951	11	O8STP4
44	502	12.2		860	11	O8CUF9
45	501.5	12.2		1055	10	O7YOO1
						O7YOO1 oryza sativ
						O8CDG1 mus musculu
						O8TC59 homo sapien
						O76B22 drosophila
						O9WME6 mus musculu
						P90T86 caenorhabd
						O9ESM6 homo sapien
						O9GPAS strongiloe
						O17567 caenorhabd
						O8ML6 dictyosteli
						O8EIF5 dictyostei
						O8IC98 stylonchia
						O8ECV7 stylyonchia
						O9NMZ8 homo sapien
						O7ZMB5 brachydanio
						O8MKZ9 tetrahymena
						O8MQC1 tetrahymena
						O9UJC9 parametium
						O8SIR2 oryza sativ
						O9SHF2 arabidopsis
						O8CTG0 mus musculu
						O9SHF3 arabidopsis
						O8LPU0 oryza sativ
						O8EBJ3 caenorhabd
						O16720 caenorhabd
						O8CPJ8 mus musculu
						O8AYI4 arabidopsis
						O8STP4 mus musculu
						O8CUF9 mus musculu
						O7YOO1 oryza sativ

ALIGNMENTS

RESULT 1

ID Q96JDS PRELIMINARY; PRT; 775 AA.

AC 33-002', 2001 (TtEMBLrel. 19, Created)
 DT 01-DEC-2001 (TtEMBLrel. 19, Created)
 DT 01-DEC-2001 (TtEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TtEMBLrel. 22, Last annotation update)
 DE HIWI.
 OS Homo sapiens (Human).
 OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Sharma A.K., Nelson M.C., Brandt J.E., Weissman M., Mahmud N.,
 RA Weller K.P., Hoffman R.;
 RT "Human CD34+ Stem Cells Express the hwi Gene, a Human Homolog of the
 RT Drosophila Gene Piv1.";
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF264004; AAK92281.1; -.
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; Piv1.
 DR Pfam; PF02171; Piv1; 1.
 DR Pfam; PF02170; PAZ; 1.
 DR PROSITE; PSS0822; PAZ; 1.
 DR PROSITE; PSS0822; Piv1; 1.
 QO SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;

	Query Match	100.0%	Score 4102	DB 4	Length 775
	Best Local Similarity	100.0%	Pred. No. 0		
	Matches 775	Conservative 0	Matches 0	Indels 0	Gaps 0
Qy	1	MIGVYTRQNLHVESKTSGGIVRLSTNHFILTSRPGWALQYHIDNPIIMEARLR	60		
Dy	1	MIGVYTRQNLHVESKTSGGIVRLSTNHFILTSRPGWALQYHIDNPIIMEARLR	60		
Zy	61	SALLPHELLIKCAFGOTILFLPKRLQGVTVFSKTRNGEYVRITITLNELPSTP	120		

```

Db      61  SALLFQHEEDLIGKCHAFDGTLLF.PKRLQOKVTEVFSEKTNGBDVRTITLTNLEPPTSP 120
Qy      121  TCLOFNILIRRLIKINMLQOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 180
Db      121  TCLOFNILIRRLIKINMLQOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 180
Qy      181  TDVSHKVLRSFTVLDPMFNPHYQTEHKKFOYSKELIGLVLT.KYNNKTYRVDIDMDQ 240
Db      181  TDVSHKVLRSFTVLDPMFNPHYQTEHKKFOYSKELIGLVLT.KYNNKTYRVDIDMDQ 240
Qy      241  NPKSTFKADGSEVSEFLEYRKQYNQEBITDLKQPVLSQPRRRPGGTLPGPAMLPEL 300
Db      241  NPKSTFKADGSEVSEFLEYRKQYNQEBITDLKQPVLSQPRRRPGGTLPGPAMLPEL 300
Qy      301  CYLGLGTDKMRNDNNMKDLAVHTRLTPEORQREVGRLIDYIHKNDNVQRELDWGLSPD 360
Db      301  CYLGLGTDKMRNDNNMKDLAVHTRLTPEORQREVGRLIDYIHKNDNVQRELDWGLSPD 360
Qy      361  SNLISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 420
Db      361  SNLISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 420
Qy      421  AANSLIQNLFKVTAPMGQMKAKAIMIEVDRETEAYLAVLQOKTADTOIYVCLLSNRKD 480
Db      421  AANSLIQNLFKVTAPMGQMKAKAIMIEVDRETEAYLAVLQOKTADTOIYVCLLSNRKD 480
Qy      481  KYDAIKKYLCTDCEPTPSQCVVARTLGKQVTMAIATKIALQNMCKMGELMRVDIPLKLV 540
Db      481  KYDAIKKYLCTDCEPTPSQCVVARTLGKQVTMAIATKIALQNMCKMGELMRVDIPLKLV 540
Qy      541  MIVGIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAM 600
Db      541  MIVGIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAM 600
Qy      601  NSCNEYMSRIIYVRDVGQGLKTLVNYEVPQFLDCLKSIGRGYNPLTVIYVKKRVNT 660
Db      601  NSCNEYMSRIIYVRDVGQGLKTLVNYEVPQFLDCLKSIGRGYNPLTVIYVKKRVNT 660
Qy      661  RFPAGSGRLQNP.LPGTYIDVEYTRPEWYDFIVSQAVRSGSVSTHYNVYDNGSLKPD 720
Db      661  RFPAGSGRLQNP.LPGTYIDVEYTRPEWYDFIVSQAVRSGSVSTHYNVYDNGSLKPD 720
Qy      721  HIQRLTYLCHIIYNNMPGIVRVPACQYAHKLAFLVGSIHREPNLSINRLYYL 775
Db      721  HIQRLTYLCHIIYNNMPGIVRVPACQYAHKLAFLVGSIHREPNLSINRLYYL 775

RESULT 2
Q96J94      PRELIMINARY;      PRT;      861 AA.
AC      096J94;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE      Homo sapiens (Human).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Sha J.H.;
RT      "Cloning and identification of human piwi protein related to testis
RT      development."
RL      Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AF87507; AK69348.1;
DR      InterPro; IPR003100; PAZ.
DR      InterPro; IPR003165; PIWI.
DR      Pfam; PF02170; PAZ; 1.
DR      Pfam; PF02171; PIWI; 1.
DR      PROSITE; PS00821; PAZ; 1.
DR      PROSITE; PS00822; PIWI; 1.

```

```

SQ      SEQUENCE      861 AA; 98603 MW; 58D76C7321DEFA4 CRC64;
Query Match      99.6%; Score 4087; DB 4; Length 861;
Best Local Similarity 100.0%; Pred. No. 1e-317;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4  GYNTRONLDHYKESKTSGGIIVRLSTNHFPLTSRPMALYQYHIDVNPMLRRLSAL 63
Db      90  GYNTRONLDHYKESKTSGGIIVRLSTNHFPLTSRPMALYQYHIDVNPMLRRLSAL 149
Qy      64  LFOHEDLIGKCHAFDGTLLF.PKRLQOKVTEVFSEKTNGBDVRTITLTNLEPPTSPCL 123
Db      150  LFOHEDLIGKCHAFDGTLLF.PKRLQOKVTEVFSEKTNGBDVRTITLTNLEPPTSPCL 209
Qy      124  QFYNIIFRLLIKINMLQOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 183
Db      210  QFYNIIFRLLIKINMLQOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 269
Qy      184  SHKVLRSFTVLDPMFNPHYQTEHKKFOYSKELIGLVLT.KYNNKTYRVDIDMDQNK 243
Db      270  SHKVLRSFTVLDPMFNPHYQTEHKKFOYSKELIGLVLT.KYNNKTYRVDIDMDQNK 329
Qy      244  STFKADGSEVSEFLEYRKQYNQEBITDLKQPVLSQPRRRPGGTLPGPAMLPEL 303
Db      330  STFKADGSEVSEFLEYRKQYNQEBITDLKQPVLSQPRRRPGGTLPGPAMLPEL 389
Qy      304  TGLTDKMRNDNNMKDLAVHTRLTPEORQREVGRLIDYIHKNDNVQRELDWGLSPD 363
Db      390  TGLTDKMRNDNNMKDLAVHTRLTPEORQREVGRLIDYIHKNDNVQRELDWGLSPD 449
Qy      364  LSFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 423
Db      450  LSFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 509
Qy      424  SLIQNLFKVTAPMGQMKAKAIMIEVDRETEAYLAVLQOKTADTOIYVCLLSNRKD 483
Db      510  SLIQNLFKVTAPMGQMKAKAIMIEVDRETEAYLAVLQOKTADTOIYVCLLSNRKD 569
Qy      484  AIKKYLCTDCEPTPSQCVVARTLGKQVTMAIATKIALQNMCKMGELMRVDIPLKLV 543
Db      570  AIKKYLCTDCEPTPSQCVVARTLGKQVTMAIATKIALQNMCKMGELMRVDIPLKLV 629
Qy      544  GIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAM 603
Db      630  GIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAM 689
Qy      604  NEYMSRIIYVRDVGQGLKTLVNYEVPQFLDCLKSIGRGYNPLTVIYVKKRVNT 663
Db      690  NEYMSRIIYVRDVGQGLKTLVNYEVPQFLDCLKSIGRGYNPLTVIYVKKRVNT 749
Qy      664  AOSGRLQNP.LPGTYIDVEYTRPEWYDFIVSQAVRSGSVSTHYNVYDNGSLKPD 723
Db      750  AOSGRLQNP.LPGTYIDVEYTRPEWYDFIVSQAVRSGSVSTHYNVYDNGSLKPD 809
Qy      724  RLTYLCHIIYNNMPGIVRVPACQYAHKLAFLVGSIHREPNLSINRLYYL 775
Db      810  RLTYLCHIIYNNMPGIVRVPACQYAHKLAFLVGSIHREPNLSINRLYYL 861

RESULT 3
Q87B5      PRELIMINARY;      PRT;      861 AA.
AC      Q87B5;
DT      01-JUN-2002 (Tremblrel. 21, Created)
DT      01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT      01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE      piwi-like 1 (Drosophila).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.

```

Wed May 12 09:50:22 2004

us-10-043-774b-2.rapt

Page 3

RC TISSUE=Testis;
RA Struhsberg R.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028581; AAH28581.1; -.
DR Genew; HGNC:9007; PIWI1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 861 AA; 98545 MW; D33376EDED743A CRC64;

Query Match 99.3%; Score 4072; DB 4; Length 861;
Best Local Similarity 99.6%; Pred. No. 1.5e-316;
Matches 769; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVNTRQNDLVKESKTSSGIIIVRLSTNHFRLTSRQWALYQHIDYNPIMEARLRSL 63
DB GVNTRQNDLVKESKTSSGIIIVRLSTNHFRLTSRQWALYQHIDYNPIMEARLRSL 149
QY 64 LFQHEHDIGKCHAFDGTLLFPRLOQKTEVSKTRNGEDVRIITLLTNELPTSPCT 123
DB LFQHEHDIGKCHAFDGTLLFPRLOQKTEVSKTRNGEDVRIITLLTNELPTSPCT 209
QY 124 QFYNIIFRLLKIMNLOQIGRNYNPNPIDIPSHRLVYWGFTTSLQYENSIMLCTDV 183
DB QFYNIIFRLLKIMNLOQIGRNYNPNPIDIPSHRLVYWGFTTSLQYENSIMLCTDV 269
QY 184 SHKVRSEITVDPMFNPHQTEHKEFOEVSKEILGLVYLTKNNKTRVDDIMDQNP 243
DB SHKVRSEITVDPMFNPHQTEHKEFOEVSKEILGLVYLTKNNKTRVDDIMDQNP 329
QY 244 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLPAPMILPELCYL 303
DB STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLPAPMILPELCYL 389
QY 304 TGLDKRNDPMVKDLAVHRLTPRQREVGRLIDYIHKNQVQRELDMWGLSFDNL 363
DB TGLDKRNDPMVKDLAVHRLTPRQREVGRLIDYIHKNQVQRELDMWGLSFDNL 449
QY 364 LPSGRILQTEKHOGKTEFDNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAN 423
DB LPSGRILQTEKHOGKTEFDNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAN 509
QY 424 SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 483
DB SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 569
QY 510 SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 569
QY 484 AIKKYLCTDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGELMRVDIPLKLVIV 543
DB AIKKYLCTDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGELMRVDIPLKLVIV 629
QY 570 AIKKYLCTDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGELMRVDIPLKLVIV 543
QY 544 GIDCYHDMTAGRRSLAGFVASINEGTRWFSRCIFQDRGQELVLDLQKLCQALALANSC 603
DB GIDCYHDMTAGRRSLAGFVASINEGTRWFSRCIFQDRGQELVLDLQKLCQALALANSC 689
QY 604 NEYMPRIIVYRDGVGDQGLKTLVNEVQPLDCLKSIRGYNPLIYIVAKRYNTRPF 663
DB NEYMPRIIVYRDGVGDQGLKTLVNEVQPLDCLKSIRGYNPLIYIVAKRYNTRPF 749
QY 664 AOSGGLQNPFGTVIDEVTPEWYDFIVSOAVRSGVSPTHNVIVYDNGSLPDIHQ 723
DB AOSGGLQNPFGTVIDEVTPEWYDFIVSOAVRSGVSPTHNVIVYDNGSLPDIHQ 809
QY 724 RLTYKLGHIYVWPGVIRVAPCOVAHKLAFIVGSGIHREPNLSISNRLYYL 775
DB RLTYKLGHIYVWPGVIRVAPCOVAHKLAFIVGSGIHREPNLSISNRLYYL 861

AC 095404;
DT 01-MAY-1999 (TRENBERG). 10, Created)
DT 01-MAY-2002 (TRENBERG). 20, Last sequence update)
DT 01-JUN-2003 (TRENBERG). 24, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Bakayota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA MEDLINE=99069219; PubMed=9851978;
RX Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
RT essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727 (1998).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
RT cell gene family whose overexpression is correlated to seminomas.";
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104260; AAC97571.2; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 861 AA; 98530 MW; 43D7F60B9D997B7 CRC64;

Query Match 99.2%; Score 4071; DB 4; Length 861;
Best Local Similarity 99.7%; Pred. No. 1.8e-316;
Matches 770; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVNTRQNDLVKESKTSSGIIIVRLSTNHFRLTSRQWALYQHIDYNPIMEARLRSL 63
DB GVNTRQNDLVKESKTSSGIIIVRLSTNHFRLTSRQWALYQHIDYNPIMEARLRSL 149
QY 64 LFQHEHDIGKCHAFDGTLLFPRLOQKTEVSKTRNGEDVRIITLLTNELPTSPCT 123
DB LFQHEHDIGKCHAFDGTLLFPRLOQKTEVSKTRNGEDVRIITLLTNELPTSPCT 209
QY 124 QFYNIIFRLLKIMNLOQIGRNYNPNPIDIPSHRLVYWGFTTSLQYENSIMLCTDV 183
DB QFYNIIFRLLKIMNLOQIGRNYNPNPIDIPSHRLVYWGFTTSLQYENSIMLCTDV 269
QY 184 SHKVRSEITVDPMFNPHQTEHKEFOEVSKEILGLVYLTKNNKTRVDDIMDQNP 243
DB SHKVRSEITVDPMFNPHQTEHKEFOEVSKEILGLVYLTKNNKTRVDDIMDQNP 329
QY 244 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLPAPMILPELCYL 303
DB STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLPAPMILPELCYL 389
QY 304 TGLDKRNDPMVKDLAVHRLTPRQREVGRLIDYIHKNQVQRELDMWGLSFDNL 363
DB TGLDKRNDPMVKDLAVHRLTPRQREVGRLIDYIHKNQVQRELDMWGLSFDNL 449
QY 364 LPSGRILQTEKHOGKTEFDNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAN 423
DB LPSGRILQTEKHOGKTEFDNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAN 509
QY 424 SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 483
DB SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 569
QY 510 SLIQNLFKYTPAMGQMRKAIMIEVDRTAVALYLOQKVTADTOIVVCLLSNRKDYD 569
QY 484 AIKKYLCTDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGELMRVDIPLKLVIV 543
DB AIKKYLCTDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGELMRVDIPLKLVIV 629

QY 544 GIDCYHMTAGRRSIAAGFVASINEGTMWFSRCIFODRGGLVDGLKVCLOAALRAMNSC 603
 DB 630 GIDCYHMTAGRRSIAAGFVASINEGTMWFSRCIFODRGGLVDGLKVCLOAALRAMNSC 689
 QY 604 NEWPSRIIYVRDVGGLKTLVNYEVPFLDCLKSIGRGYNPLTVIVKRVNTRFF 663
 DB 690 NEWPSRIIYVRDVGGLKTLVNYEVPFLDCLKSIGRGYNPLTVIVKRVNTRFF 749
 QY 664 AOSGGRLONPGLPGVIVDEVTREPMYDFFTVSOAVRSGSVSPTHYNYIDNSGLKPDHIO 723
 DB 750 AOSGGRLONPGLPGVIVDEVTREPMYDFFTVSOAVRSGSVSPTHYNYIDNSGLKPDHIO 809
 QY 724 RLTYKLCHYIYNNPGVIRVAPPCQYAHKLAFLVGSIRREPNLSLSNRLYYL 775
 DB 810 RLTYKLCHYIYNNPGVIRVAPPCQYAHKLAFLVGSIRREPNLSLSNRLYYL 861

RESULT 5

Q9JMB7 PRELIMINARY; PRT; 862 AA.

AC 09JMB7
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE MIWI (P1wi).
 GN PIWI1 OR MIWI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Miyagawa S.K., Kimura T., Nakano T.;
 RT "Molecular Cloning and characterization of P1wi family genes."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Deng W., Lin H.;
 RT "Mwi, a murine homolog of P1wi, encodes a cytoplasmic protein
 essential for spermatogenesis."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB032604; BAA33705.1; -
 DR EMBL; AF438405; AAL31014.1; -
 DR MGD; MGI:1928897; P1wi1.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0003727; F:single-stranded RNA binding; IDA.
 DR GO; GO:0007283; P:spermatogenesis; IMP.
 DR InterPro; IPR003100; P1wi.
 DR InterPro; IPR003165; P1wi.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1wi; 1.
 DR PROSITE; PSS0821; PAZ; 1.
 DR PROSITE; PSS0822; P1wi; 1.
 SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CC04C CRC64;

Query Match 97.0%; Score 3980; DB 11; Length 862;
 Best Local Similarity 96.4%; Pred. No. 3,4e-309;
 Matches 744; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 4 GNTFONLDHYVESTGSSGIIIVRLSTNHFRILTSRPPALQYHIDYVPLMEARRLSAL 63
 DB 91 GNTFONLDHYVESTGSSGIIIVRLSTNHFRILTSRPPALQYHIDYVPLMEARRLSAL 150
 QY 64 LFQHEDLIGKCHAFDGTLLFLPKRLQQVTEVFSKTRNGEDVRIITITLNELPPTSPCL 123
 DB 151 LFQHEDLIGKCHAFDGTLLFLPKRLQQVTEVFSKTRNGEDVRIITITLNELPPTSPCL 210
 QY 124 QFYNIIFRRLKIKMLQOIGRYVYVNPDPIDPSRLVIYVPGFTSIIQYENSIMLCTDV 183
 DB 211 QFYNIIFRRLKIKMLQOIGRYVYVNPDPIDPSRLVIYVPGFTSIIQYENSIMLCTDV 270
 QY 184 SHKVLRSFETVLDPMENFYHQTEHFKFOQSVSKELGLVVLTKYNNKTYRVDDIDMDONPK 243

DB 271 SHKVLRSFETVLDPMENFYHQTEHFKFOQSVSKELGLVVLTKYNNKTYRVDDIDMDONPK 330
 QY 244 STFKKADSEVSEFLEYRKQYNOEITDLKQPLVSOQRRRRGPGGTLPGPMLIPELCYL 303
 DB 331 STFKKADSEVSEFLEYRKQYNOEITDLKQPLVSOQRRRRGPGGTLPGPMLIPELCYL 390
 QY 304 TGLTPKANDPNVWKDLAVHRLTPEQORPEVGLIYIHKNIDVORELSDMGSPDNL 363
 DB 391 TGLTPKANDPNVWKDLAVHRLTPEQORPEVGLIYIHKNIDVORELSDMGSPDNL 450
 QY 364 LSFSGRLIQTEKIHGGKTFPDYNPOFADWSKETGAPLISVYPLDMLIYTRNRYEAA 423
 DB 451 LSFSGRLIQSEKIHGGKTFPDYNPOFADWSKETGAPLISVYPLDMLIYTRNRYEAA 510
 QY 424 SLIQMLFVTPAMGMQKAIMIEVDRTETAYLRLQOKTADTQIVVCLSSNRKQYD 483
 DB 511 SLIQMLFVTPAMGMQKAIMIEVDRTETAYLRLQOKTADTQIVVCLSSNRKQYD 570
 QY 484 AIKKYLCTDCEPTSPQCVVARTGKQOYMAIATKIALQMNCKGSELRVDIPLKLYIV 543
 DB 571 AIKKYLCTDCEPTSPQCVVARTGKQOYMAIATKIALQMNCKGSELRVDIPLKLYIV 630
 QY 544 GIDCYHMTAGRRSIAAGFVASINEGTMWFSRCIFODRGGLVDGLKVCLOAALRAMNSC 603
 DB 631 GIDCYHMTAGRRSIAAGFVASINEGTMWFSRCIFODRGGLVDGLKVCLOAALRAMNSC 690
 QY 604 NEWPSRIIYVRDVGGLKTLVNYEVPFLDCLKSIGRGYNPLTVIVKRVNTRFF 663
 DB 691 NEWPSRIIYVRDVGGLKTLVNYEVPFLDCLKSIGRGYNPLTVIVKRVNTRFF 750
 QY 664 AOSGGRLONPGLPGVIVDEVTREPMYDFFTVSOAVRSGSVSPTHYNYIDNSGLKPDHIO 723
 DB 751 AOSGGRLONPGLPGVIVDEVTREPMYDFFTVSOAVRSGSVSPTHYNYIDNSGLKPDHIO 810
 QY 724 RLTYKLCHYIYNNPGVIRVAPPCQYAHKLAFLVGSIRREPNLSLSNRLYYL 775
 DB 811 RLTYKLCHYIYNNPGVIRVAPPCQYAHKLAFLVGSIRREPNLSLSNRLYYL 862

RESULT 6

Q9NAG0 PRELIMINARY; PRT; 829 AA.

AC 09NAG0
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical protein FLN35814 (P1wi).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Testis;
 RA Oshima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,
 RA Arima M., Masehino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Magatsuma M.,
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isegai T.;
 RT "NED0 human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK093133; BAC04068.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; P1wi.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1wi; 1.
 DR PROSITE; PSS0821; PAZ; 1.
 DR PROSITE; PSS0822; P1wi; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 829 AA; 94816 MW; 3AF9FD1D0FA9F5A7 CRC64;

Wed May 12 09:50:22 2004

us-10-043-774b-2.rpt

Page 5

Query Match
Best Local Similarity 94.7%; Score 3883; DB 4; Length 829;
Matches 733; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GNTNQNDLHVKESTKSGSGLIVRLSTNHFRLTSRPPOMALYQYHIDYPLMEARLSAL 63
DB GNTNQNDLHVKESTKSGSGLIVRLSTNHFRLTSRPPOMALYQYHIDYPLMEARLSAL 149
QY 64 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTNLEPPTPTCL 123
DB LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTNLEPPTPTCL 209
QY 124 QFYNIIFRRLKTNMLOOIGRNYNPNPIDPSHRLVIMPGFTTSLIQYENSIMLCTDV 183
DB QFYNIIFRRLKTNMLOOIGRNYNPNPIDPSHRLVIMPGFTTSLIQYENSIMLCTDV 269
QY 184 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVLTFRKNNKTYRVDIDMDONPK 243
DB SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVLTFRKNNKTYRVDIDMDONPK 329
QY 244 STFKADGSEVSEFLYYRKYNOEITDLKQPVLSQPKRRRGPGGTLDPGAMLIPELCYL 303
DB STFKADGSEVSEFLYYRKYNOEITDLKQPVLSQPKRRRGPGGTLDPGAMLIPELCYL 389
QY 304 TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 363
DB TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 449
QY 390 TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 449
DB TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 509
QY 424 SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 483
DB SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 569
QY 510 SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 569
DB SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 629
QY 544 GIDCYHDMTAGRRSIAFGVASINSGMTRFSCRIFODRGQELVDGLKVCQALPAMNSC 603
DB GIDCYHDMTAGRRSIAFGVASINSGMTRFSCRIFODRGQELVDGLKVCQALPAMNSC 689
QY 604 NEYMSRIIVYRDGVGDQGLKTLVNYEVPQFLDKSISGRGYNPLTVIYVKKRYNTRFF 663
DB NEYMSRIIVYRDGVGDQGLKTLVNYEVPQFLDKSISGRGYNPLTVIYVKKRYNTRFF 749
QY 664 AOSGRLQNPPLPGVIVDEVTRPEMYDFFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 723
DB AOSGRLQNPPLPGVIVDEVTRPEMYDFFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 809
QY 724 RLTYKLCMTYNNWP 737
DB RLTYKLCMTYNNWP 823
RESULT 7
Q8UYX0 PRELIMINARY; PRT; 858 AA.
AC 08UYX0; PRELIMINARY; PRT; 858 AA.
DT 01-MAR-2002 (TREMELREL. 20, Created)
DT 01-MAR-2002 (TREMELREL. 20, Last sequence update)
DT 01-OCT-2002 (TREMELREL. 22, Last annotation update)
DE Plw1 protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.

RA Weeratane S.D., Gong Z., Tan C.-H.;
RT "Cloning and characterization of zebrafish homolog of plw1, essential
for germ-line stem cell self-renewal."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF36369; AL57170.1; -
DR InterPro: IPR003100; PAZ;
DR InterPro: IPR003165; Plw1.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Plw1; 1.
DR PROSITE: PS50821; PAZ; 1.
DR PROSITE: PS50822; Plw1; 1.
SQ SEQUENCE 858 AA; 97451 MW; 6A12F2511465777 CRC64;

Query Match
Best Local Similarity 69.0%; Score 2828.5; DB 13; Length 858;
Matches 509; Conservative 121; Mismatches 141; Indels 1; Gaps 1;

QY 4 GNTNQNDLHVKESTKSGSGLIVRLSTNHFRLTSRPPOMALYQYHIDYPLMEARLSAL 63
DB GNTNQNDLHVKESTKSGSGLIVRLSTNHFRLTSRPPOMALYQYHIDYPLMEARLSAL 147
QY 64 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTNLEPPTPTCL 123
DB LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTNLEPPTPTCL 207
QY 124 QFYNIIFRRLKTNMLOOIGRNYNPNPIDPSHRLVIMPGFTTSLIQYENSIMLCTDV 183
DB QFYNIIFRRLKTNMLOOIGRNYNPNPIDPSHRLVIMPGFTTSLIQYENSIMLCTDV 267
QY 184 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVLTFRKNNKTYRVDIDMDONPK 243
DB SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVLTFRKNNKTYRVDIDMDONPK 327
QY 244 STFKADGSEVSEFLYYRKYNOEITDLKQPVLSQPKRRRGPGGTLDPGAMLIPELCYL 303
DB STFKADGSEVSEFLYYRKYNOEITDLKQPVLSQPKRRRGPGGTLDPGAMLIPELCYL 386
QY 304 TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 363
DB TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 446
QY 390 TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 446
DB TGLTTRKNDPMVMDLAVHTLRLTEPQORQREYGRILIDYTHKNDVQRLRPMGLSPDSNL 506
QY 424 SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 483
DB SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 566
QY 510 SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 566
DB SLIQNLFKYTPAMQMKRAIMIEVDRTAVRLVLOQKVTADTOIVCLLSNRKDXD 629
QY 544 GIDCYHDMTAGRRSIAFGVASINSGMTRFSCRIFODRGQELVDGLKVCQALPAMNSC 603
DB GIDCYHDMTAGRRSIAFGVASINSGMTRFSCRIFODRGQELVDGLKVCQALPAMNSC 686
QY 604 NEYMSRIIVYRDGVGDQGLKTLVNYEVPQFLDKSISGRGYNPLTVIYVKKRYNTRFF 663
DB NEYMSRIIVYRDGVGDQGLKTLVNYEVPQFLDKSISGRGYNPLTVIYVKKRYNTRFF 746
QY 664 AOSGRLQNPPLPGVIVDEVTRPEMYDFFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 723
DB AOSGRLQNPPLPGVIVDEVTRPEMYDFFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 806
QY 724 RLTYKLCMTYNNWP 737
DB RLTYKLCMTYNNWP 807
RESULT 8
Q72323 PRELIMINARY; PRT; 882 AA.

AC Q7223; 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE PIVIL3 protein.
 GN PIVIL3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22788786; PubMed=12906857;
 RA Sasaki T., Shiohama A., Minoshima S., Shimizu N.,
 RT "Identification of eight members of the Argonaute family in the human
 genome.";
 RL Genomics 82:323-330(2003).
 DR EMBL, AB079368; BAC01343.1; -
 SQ SEQUENCE 882 AA; 101121 MW; BD65CA74C3D36DB8 CRC64;

Query Match 52.3%; Score 2144; DB 4; Length 882;
 Best Local Similarity 52.0%; Pred. No. 2.6e-162;
 Matches 407; Conservative 139; Mismatches 222; Indels 14; Gaps 5;

QY 5 VNTKRONLDHVRESKTSKSSGIIIVRLSTNFRILTSRPMALYOHYHIDNPMLEARRLSALL 64
 DB 104 VNTKRONLDHVRESKTSKSSGIIIVRLSTNFRILTSRPMALYOHYHIDNPMLEARRLSALL 163
 QY 65 FCHEDLIGKCAFDGTLIFLPRKLOQKTEVFSTRNEDVRIITITNELPPTSPCLQ 124
 DB 164 DQHRKRPERRHIFDONSLLSRPLKERVLELSTTKDNVYKIVERSKELTSPDCLR 223
 QY 125 FYNIFRRLKIMNIQIGRNYYNPNNDIDIPSH-RUYVPGTTSILQYENSIMCTD 182
 DB 224 YNIFRFRRLKIMNIQIGRNYYNPNNDIDIPSH-RUYVPGTTSILQYENSIMCTD 283
 QY 183 VSHKYLRSSETVLDENFNYHOTEHKEFOEVSKEHIGLVLTATKNNKYRVDIDMDQNP 242
 DB 284 VSHKYLRSSETVLDENFNYHOTEHKEFOEVSKEHIGLVLTATKNNKYRVDIDMDQNP 343
 QY 243 KSTFKKADGSEVSLEYRYKQYNOETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPELCY 302
 DB 344 ETEFKSDGSKITVYIDYRQOKEIVYKOPPLVYSCGRWKKGTLGTQREPIILLIPOLCH 403
 QY 303 LTGLTDXKRNPNWMDLAHTRLPBEQOREVGLIDYHKNNVORELSDWGLSPDSN 362
 DB 404 MTGLTDXKRNPNWMDLAHTRLPBEQOREVGLIDYHKNNVORELSDWGLSPDSN 463
 QY 463 LLSFSGRLIQTEKIHOGKTFDYNPOFADWSKETRGAPLISVKPLDNMLIYTRRYEAA 422
 DB 522 FLISVGRVILKNNIYQGRRWYKNSQ-GDWSREIREFLNAMPFLSHWLIYSSRSREA 522
 QY 423 NSLIONLFKVTYPAMGOKRKAIMEVDRTAYLRVLOQKTYADLQI-----VYCL 473
 DB 522 NSLIONLFKVTYPAMGOKRKAIMEVDRTAYLRVLOQKTYADLQI-----VYCL 582
 QY 474 LLSNRKDKYDAIKKYLCTDCPTPSQCVARTLQKQYMAIATKIALONCMKMGELMRV 533
 DB 582 LLSNRKDKYDAIKKYLCTDCPTPSQCVARTLQKQYMAIATKIALONCMKMGELMRV 641
 QY 534 DIPKLVLVIGIDCYHDMTAGRSIAGYASINEGMTWFSRCIFQDNGSELINDGAKYL 593
 DB 642 DIPKLVLVIGIDCYHDMTAGRSIAGYASINEGMTWFSRCIFQDNGSELINDGAKYL 701
 QY 594 QALDAMNSCNEFMSRIIVYRDGVDDGQLKTLVAYEVPOFLDCKSIGRGVNPFLVIV 653
 DB 702 QALDAMNSCNEFMSRIIVYRDGVDDGQLKTLVAYEVPOFLDCKSIGRGVNPFLVIV 760
 QY 654 VKKRVNTRFPAQSGRLQNPFLPGVIVDVEVTRPEMYDFFIYSGAVRSGSVSPTHNVYD 713
 DB 761 VKKRVNTRFPAQSGRLQNPFLPGVIVDVEVTRPEMYDFFIYSGAVRSGSVSPTHNVYD 820
 QY 714 NSGLKPDHQLRLTYLCHLYNMPGIVRVPACQYAHKLAFLVQSGISREPIILSNRLY 773

DB 821 TIGLSPDVTQRLTYLCHLYNMPGIVRVPACQYAHKLAFLVQSGISREPIILSNRLY 880
 QY 774 YL 775
 DB 881 YL 882

RESULT 9
 ID 08NEH2 PRELIMINARY; PRT; 852 AA.
 AC 08NEH2;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to p1w1 like homolog 1 (Drosophila).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, BC031060; AAH31060.1; -
 DR InterPro: IPR003100; PAF;
 DR InterPro: IPR003165; P1w1.
 DR Pfam: PF02170; PAF; 1.
 DR Pfam: PF02170; P1w1; 1.
 DR PROSITE: PS00821; PAF; 1.
 DR PROSITE: PS00822; P1w1; 1.
 SQ SEQUENCE 852 AA; 96588 MW; 37769BE078B96D13 CRC64;

Query Match 51.3%; Score 2103.5; DB 4; Length 852;
 Best Local Similarity 51.2%; Pred. No. 4.2e-159;
 Matches 398; Conservative 146; Mismatches 225; Indels 9; Gaps 4;

QY 1 MFGVNTKRONLDHVRESKTSKSSGIIIVRLSTNFRILTSRPMALYOHYHIDNPMLEARRLSALL 60
 DB 81 MFGVNTKRONLDHVRESKTSKSSGIIIVRLSTNFRILTSRPMALYOHYHIDNPMLEARRLSALL 140
 QY 61 SALLFQHEDLIGKCAFDGTLIFLPRKLOQKTEVFSTRNEDVRIITITNELPPTSPCLQ 120
 DB 141 SALLFQHEDLIGKCAFDGTLIFLPRKLOQKTEVFSTRNEDVRIITITNELPPTSPCLQ 200
 QY 121 TCIQFNIIIFRRLKIMNIQIGRNYYNPNNDIDIPSHRVLVMPGTTTSLQYENSIMCTD 180
 DB 201 TCIQFNIIIFRRLKIMNIQIGRNYYNPNNDIDIPSHRVLVMPGTTTSLQYENSIMCTD 260
 QY 181 TDVSHKYLRSSETVLDENFNYHOTEHKEFOEVSKEHIGLVLTATKNNKYRVDIDMDQNP 240
 DB 261 TDVSHKYLRSSETVLDENFNYHOTEHKEFOEVSKEHIGLVLTATKNNKYRVDIDMDQNP 320
 QY 241 NPSTFKKADGSEVSLEYRYKQYNOETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPELCY 300
 DB 321 NPSTFKKADGSEVSLEYRYKQYNOETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPELCY 379
 QY 301 CYYLGLTDXKRNPNWMDLAHTRLPBEQOREVGLIDYHKNNVORELSDWGLSPDSN 360
 DB 380 CYYLGLTDXKRNPNWMDLAHTRLPBEQOREVGLIDYHKNNVORELSDWGLSPDSN 439
 QY 361 SALLFQHEDLIGKCAFDGTLIFLPRKLOQKTEVFSTRNEDVRIITITNELPPTSPCLQ 417
 DB 440 SALLFQHEDLIGKCAFDGTLIFLPRKLOQKTEVFSTRNEDVRIITITNELPPTSPCLQ 494
 QY 418 NYEANSLLIONLFKVTYPAMGOKRKAIMEVDRTAYLRVLOQKTYADLQI-----VYCL 477
 DB 495 NYEANSLLIONLFKVTYPAMGOKRKAIMEVDRTAYLRVLOQKTYADLQI-----VYCL 554
 QY 478 RKKYDAIKKYLCTDCPTPSQCVARTLQKQYMAIATKIALONCMKMGELMRV 537
 DB 555 RKKYDAIKKYLCTDCPTPSQCVARTLQKQYMAIATKIALONCMKMGELMRV 614

QY 538 KLVMIIGIDCYHDMTAGRSIAGFVASINEGNTWFSRCIFQDRQGLVDGLKYCLOAL 597
 DB 615 KSLMVGIDVCKDALSCKVMVGVASVNPRIITWFSRCILQRTWTDVADCLKFMTCAL 674
 QY 598 RAMNSCNEYMPRIIYRDGVGDQGLKTLVNEVPOFLDCKLSIGRGVNPRLTVIVKRR 657
 DB 675 NKMYNNHDLPAKILYRAGVGDQGLKTLIEYVPOFLDSVASSSNTSRSLVIVKRR 734
 QY 658 VNTREFAOSGGRLONPFGTIVDEVTREPWTDFEIVSQAVRSGVSPTHYNYTNSGL 717
 DB 735 CMPEFTEKMRITVQNPPIGLTVDSBATRNEMVDYFLISQVACRGTVSPITYNYIYDNGL 794
 QY 718 KPDHQRILTYKLCIYYNMPGVIRPAPCOYAHKLAFLVGSISIREPNTLSNRLTYL 775
 DB 795 KPDHQRILTYKLCIYYNMPGVIRPAPCOYAHKLAFLVGSISIREPNTLSNRLTYL 852

RESULT 10

Q8N9V8 PRELIMINARY; PRT: 852 AA.

AC Q8N9V8; 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ36156 (p1w1).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Testis;
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Hatakeyama T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isoco Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sakine M., Kikuchi H., Kanda K.,
 RA Magatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK093475; BAC04179.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; P1w1.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1w1; 1.
 DR PROSITE; PS00821; PAZ; 1.
 DR PROSITE; PS00822; P1w1; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 852 AA; 96560 MW; 75C6FEFAE870701B CRC64;

Query Match 51.2%; Score 2101.5; DB 4; Length 852;
 Best Local Similarity 51.0%; Pred. No. 61e-159;
 Matches 397; Conservative 148; Mismatches 224; Indels 9; Gaps 4;

QY 1 MIFGVNTRONLDHVESKTSKGGSIIVRLSTNHRFLTSRPOALYOHIDYINPLMEARRR 60
 DB 81 MDLSICTREKLAHVNRCKAGSSGIPVKLVTLNFDPODQOLYOHYTVIYDPLASRRR 140
 QY 61 SALLFOHEDLIGKCHAFDGTILFLPKRLOQKYTEVSKTRNGEDVYITITLNLPTSP 120
 DB 141 IALLVSHSELSENKAKAFDGAIFLSQKLEKYTELSSETQGETTKMTITLKRLEPSSSP 200
 QY 121 TCLQFYNIIFRLKIMNLQOIGRNYNPNPIDIPSHRLVIMPGFTTSLQYENSIMLC 180
 DB 201 VCIOGFNIIFRLKIMNLQOIGRNYNPNPIDIPSHRLVIMPGFTTSLQYENSIMLC 260
 QY 181 TDVSHKVLRSFTVDFMNFYHQTREKHFQEQVSKELIGLVYLTIRANKYTRVVDIDWDQ 240
 DB 261 ADVSHKVLRSFTVDFMNFYHQTREKHFQEQVSKELIGLVYLTIRANKYTRVVDIDWDQ 320

QY 241 NPKSTFKKADGSFVSLEYRYKQYNOETIDLPKVLVSPKRRRGCGTLPGRAMTIPEL 300
 DB 321 KPTHFQKDGDTETIYVDDYKQYDITVSDINGPMLVSLKKRRDNSEAO-LAHIIPEL 379
 QY 301 CYLTGLTDKRNDFENVKMDLAVHTRLTPSROREVRGLIDYIHKNDVORELMDGLSPD 360
 DB 380 CFLTGLTDQRTSPQMLKAVAEKTRLSPSGRQRLARLDNORTNNAFETETWGLHFG 439
 QY 361 SNLSPSGRILQTEKIHGGKTFDYNPQ--PADMSKETRGAPLISVKPLDMLIYTRR 417
 DB 440 SQ-ISTLGRVPSSEKILMO---DHICQVSAADWSKDRCTCKILNAQSLMTWLILCSDR 494
 QY 418 NYEANSILQNFKTYPAMGMOKRAIMTEVDRTAVALRYVQOKTADTOIVVCLSSN 477
 DB 495 TEYVASFNLCLFRVAGSMGFNDYPKIKVGENPAAFYRAIQYVDPDQVLMCLPSN 554
 QY 478 RKDYDAIKKYLCTDPTSPQCVVARTLGGQOTVMAIATKIALQNMCKMGGELMRVDPL 537
 DB 555 OKTYYSIKKYLSDDCPVPSQCVLARTLKKQGMMSIATKIAMQMTCKLGGELMAVEIPL 614
 QY 538 KLVMIIGIDCYHDMTAGRSIAGFVASINEGNTWFSRCIFQDRQGLVDGLKYCLOAL 597
 DB 615 KSLMVGIDVCKDALSCKVMVGVASVNPRIITWFSRCILQRTWTDVADCLKFMTCAL 674
 QY 598 RAMNSCNEYMPRIIYRDGVGDQGLKTLVNEVPOFLDCKLSIGRGVNPRLTVIVKRR 657
 DB 675 NKMYNNHDLPAKILYRAGVGDQGLKTLIEYVPOFLDSVASSSNTSRSLVIVKRR 734
 QY 658 VNTREFAOSGGRLONPFGTIVDEVTREPWTDFEIVSQAVRSGVSPTHYNYTNSGL 717
 DB 735 CMPEFTEKMRITVQNPPIGLTVDSBATRNEMVDYFLISQVACRGTVSPITYNYIYDNGL 794
 QY 718 KPDHQRILTYKLCIYYNMPGVIRPAPCOYAHKLAFLVGSISIREPNTLSNRLTYL 775
 DB 795 KPDHQRILTYKLCIYYNMPGVIRPAPCOYAHKLAFLVGSISIREPNTLSNRLTYL 852

RESULT 11

Q72324 PRELIMINARY; PRT: 852 AA.

AC Q72324; 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE P1W1L1 protein.
 GN P1W1L1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22788786; PubMed=12906857;
 RA Sasaki T., Shiohama A., Minoshima S., Shimizu N.;
 RT "Identification of eight members of the Argonaute family in the human
 genome." 82:323-330 (2003).
 RL EMBL; AB079366; BAC81341.1; -
 SQ SEQUENCE 852 AA; 96558 MW; 3777A264A21AD56 CRC64;

Query Match 51.2%; Score 2098.5; DB 4; Length 852;
 Best Local Similarity 51.0%; Pred. No. 1.1e-158;
 Matches 397; Conservative 146; Mismatches 226; Indels 9; Gaps 4;

QY 1 MIFGVNTRONLDHVESKTSKGGSIIVRLSTNHRFLTSRPOALYOHIDYINPLMEARRR 60
 DB 81 MDLSICTREKLAHVNRCKAGSSGIPVKLVTLNFDPODQOLYOHYTVIYDPLASRRR 140
 QY 61 SALLFOHEDLIGKCHAFDGTILFLPKRLOQKYTEVSKTRNGEDVYITITLNLPTSP 120
 DB 141 IALLVSHSELSENKAKAFDGAIFLSQKLEKYTELSSETQGETTKMTITLKRLEPSSSP 200
 QY 121 TCLQFYNIIFRLKIMNLQOIGRNYNPNPIDIPSHRLVIMPGFTTSLQYENSIMLC 180

"The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis.";
 RT Genes Dev. 16:2733-2742 (2002).

RT [2]
 RN SEQUENCE FROM N.A.
 RA Adams M., Mural R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY155682; AAN75583.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; P1w1.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1w1; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; P1w1; 1.
 SQ SEQUENCE 878 AA; 98617 MW; 3C6265E0E4ED7C3F CRC64;

Query Match 45.3%; Score 1859.5; DB 11; Length 878;
 Best Local Similarity 46.5%; Pred. No. 1.5e-139;
 Matches 370; Conservative 134; Mismatches 209; Indels 83; Gaps 10;

QY 4 GVNTRONLDHYKEKTSSTGIIIVRLSTNHPRLSRPQWALYQYHIDYNPIMLEARRLSAL 63
 DB 142 GVTREKLTTHVXDCTGSSGIPVRLVTNLFMLDLPQDMQLYQHIVYSPDLASRRRLAL 201
 QY 64 LFOHEDLIGKCHAPDGTLLPLPKRLQOKVTEVPSKTRNGEDVRIITITLTNELPTSPCL 123
 DB 202 LYNHSLISDKAKAFDGLFLSEKLDQKVELTSETQGETIKITLTLSKLFENSPVCI 261
 QY 124 QFNVTIRRLKIKMLQOIGRNYNPNPDIDIPSHRLVIMGFTSLTLOENSMICTDV 183
 DB 262 QFNVTIRRLKIKMLQOIGRNYNPNPDIDIPSHRLVIMGFTSLTLOENSMICTDV 183
 QY 184 SHKVLRSYVDLDFMNFYHQTEHKFOYVSKELIGLVLLTKNNKTYRVDIDMDQPK 243
 DB 306 NYKTLRNETVDPTDLCRTGSCFTEMCHKOLVGLVLTFRNNKTYRIDIDMSVKT 365
 QY 244 STPKADGSEVSELYRKQYNOETITDKQPVLSQPRRRGPGGTLPPAMLIPELCTV 303
 DB 366 QAFQKRGSEVLYVYQYQYDITLSDLNQPVLSLTKRKNDNSE-PQWVHLMPCL 424
 QY 304 TGLDKNRNDPNWMDLAVHTRLPBQREVRGLIDYIHNDVQRELD-----W-- 355
 DB 425 TGLSQATSDPRMLKAVAEETRLSPVQOQALAVDIDQRTLPSSQVLSHTSLPMAP 484
 QY 356 ---GLSPD---SNLSPSGRILOTEKIHOG-----GKTPDYNPQF-ADMSKETEGA 399
 DB 485 EPGGLSSAIPSTVLPFAQQLTLALSLSPGIPLPHLKPSPFLFCQPAFAADMSKDWSC 544
 QY 400 PLISVPELDNMLLYTRNRYEANSLSLONLFKTPAMQMOKRAIMIEVDRTAAYRLV 459
 DB 545 KVLSSQPLNRLIYCCNAEHLIEAFSLCRRVSGSGENGVKTIKVDTPAFRAAI 604
 QY 460 QOKYADTQIVVCLLSNRKDKYDAIKKYLCTDPTSPSCQVAVATLKGQOTVAATKIA 519
 DB 605 QVHGPDPQVLMCLIPSNQKYVDSIKKYLSDCFVPSQCVLTGLTKNGQMLSVATKIA 664
 QY 520 LQWCKRKGELMRVDIPKLVMYIGIDCYHMTAGRSIAGFVASINEGTRMRSRCIFQ 579
 DB 665 MQMCKLGGELMSVEIPLKSLMVVGIDICRDALKNVYVGFVASINSRITRMSRCVLQ 724
 QY 580 DRGELVGLKYCQALRAWNSCNEYPSRIIYRDGVDGQQLTVNTEVPOFLDCLX 639
 DB 725 RTADIDACLVCMTCALNRYRHHNDIPARIIVYRQGVGQKAVLEVEVPGLKSVT 784
 QY 640 SIGRCINRLTYIVYKRVNTRFPAQSGRLQNPFGVIDVETRYREMTDFIVSQAVR 699
 DB 785 ECG-----SDAR-----DFTYIISQIAN 802
 QY 700 SGSVPTHYNYIYDNGGLKPDHIOQLTYLCHIIYNNPFGVIRVPAQCYAKLAFVLGQS 759
 DB 803 RGTYSPTHYNYIYDNGGLKPDHIOQLTYLCHIIYNNPFGVIRVPAQCYAKLAFVLGQS 862
 QY 760 IHREPNLSLNRLLYTL 775

DB 863 VAKPSLELANLRYL 878

RESULT 14
 ID 08N8G9 PRELIMINARY; PRT; 666 AA.
 AC 08N8G9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein Flj39518 (p1w1).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_Taxid=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
 RA Takebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Wagaatsuma M., Murakawa K., Kanehori K., Takehashi-Fuji A., Oshima A.,
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
 RA Masuko Y., Nagai K., Isogai T.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK096837; BAC04873.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; P1w1.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; P1w1; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; P1w1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 666 AA; 76039 MW; 1082D86916390FA8 CRC64;

Query Match 44.8%; Score 1838.5; DB 4; Length 666;
 Best Local Similarity 51.5%; Pred. No. 4.5e-138;
 Matches 346; Conservative 127; Mismatches 190; Indels 9; Gaps 4;

QY 107 ITTTLNLPPTSPCTQCPYNIIFPRLLKIMLQOIGRNYNPNPDIDIPSHRLVIMPGF 166
 DB 1 MTTLKRELPSSTSPICQYFNIIIFRKILKLSMYQIGRNFYPSPEMELPQKLSIMPGF 60
 QY 167 TTSILOENSMICTDVSHKVLRSYVDLDFMNFYHQTEHKFOYVSKELIGLVLLTKY 226
 DB 61 AIVSYFEKKLFPSADVSKYVLRNETHVLEFMALCQRTGLSCFTQCEQQLGLVLTREY 120
 QY 227 NNTKYRVDIDMDQPKSTFKKADGSEVSELYRKQYNOETITDKQPVLSQPRRRGPGGTL 286
 DB 121 NNTKYSIDIDMSVKTFTFLKRDGETELTYVYQYQYDITLSDLNQPVLSLTKRKNDNSE 180
 QY 287 GGLTPPAMLIPELCTVLTGLTDKMRNDFNWMKD LAVHTRLPBQREVRGLIDYIHNDV 346
 DB 181 NSE-PQLAHLIELCLTGLTQATSDFQMLKAVAEETRLSPVQOQALAVDIDQRTLPSSQV 239
 QY 347 NQRELRDMLGSPDSNLSFSGRILOTEKIHOGKTFDYNPQ---FADMSKETEGAFLIS 403
 DB 240 NARFELTEGLHFGSQ-ISLTGRIVSEKILMQ---DHICQPVSAADMSKIRTKIILN 294
 QY 404 VPELDNMLLYTRNRYEANSLSLONLFKTPAMQMOKRAIMIEVDRTAAYRLVLOQYV 463
 DB 295 AQSINMTLLCGDRIEYVAVESFLNCLARVYTGSMGVNDVPKLIIVQENPAFAAFRAIQYV 354
 QY 464 TADTQIVVCLLSNRKDKYDAIKKYLCTDPTSPSCQVAVATLKGQOTVAATKIALOXN 523
 DB 355 DDPVQVLMCLIPSNQKYVDSIKKYLSDCFVPSQCVLTGLTKNGQMLSVATKIAIMQNT 414
 QY 524 CKRKGELMRVDIPKLVMYIGIDCYHMTAGRSIAGFVASINEGTRMRSRCIFQDRQ 583
 DB 415 CILGSELMAVEIPLKSLMVVGIDVCDALSKQVMYVGVCAVNPRTTRFMSRCILQRTWT 474

Wed May 12 09:50:22 2004

us-10-043-774b-2.rsp

Page 10

```
QY 584 ELVDGKVCLOALBANS CNEYMPRIIVYRDVGQGLKTLVNEVEPQLDCLKSIGR 643
DB 475 DVADCLKVFMTGALNKYKKNHDLPAIIYRAGVGBGGLKTLIEVEVQLSSVASSS 534
QY 644 GINDELTVIYVKKRVNTFFAOGGRLONPLGCTVIDEYTRPEMTDFIVSQAVSSGV 703
DB 535 NTSRLSLVIVYKKCMERFTEKMRIVONPLGIVDSEKTRNEMWDFYILISQVACRGTV 594
QY 704 SPTNYIYDNGSLKPHIORLFTYLCIYYNMPGVIRVAPQCYAHKLAFLVQSIHRE 763
DB 595 SPTNYIYDNGSLKPHIORLFTYLCIYYNMPGVIRVAPQCYAHKLAFLVQSIHRE 654
QY 764 PNLISLNRLLYL 775
DB 655 PSLSLANHLFYL 666

RESULT 15
QYGPAT7 PRELIMINARY; PRT; 808 AA.
AC QYGPAT7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Seawi (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinioidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
CX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT Homolog of P1w1."
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY014900; AAG42534.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; P1w1.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; P1w1; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; P1w1; 1.
DR NON TER 808
SQ SEQUENCE 808 AA; 91366 MM; 19528F4B9D10474C CRC64;

Query Match 44.2%; Score 1811.5; DB 5; Length 808;
Best Local Similarity 49.4%; Pred. No. 8.9e-136;
Matches 354; Conservative 125; Mismatches 228; Indels 9; Gaps 6;

QY 14 VKESSTSSGIVLSTNHFRLTSRPOVALQYHIDVNPMLFEARLRSAALLFQHEDLIGK 73
DB 98 VKQALAGK---IALIANGFKLTKPQOLQYRVDFEPELINPRAFAIKGHSALLGK 154
QY 74 CHAFDGTLLPFLKRLQKQTEVFSKTRNGEDVRITITLTNELPPTSCLQFYNIIFRL 133
DB 155 GLTLDMDTMYSLFKLAEKVTNLSARKDGSNNICVTIVATINPMAPNTLHLVNLFFRC 214
QY 134 LKIMNLQOIGRNYNPNPIDIPSHRLVIMPFTTSIIQVENSIMLCQDVSHKYLRSSTV 193
DB 215 LKIMNEQGRNYDPTAIDIKQGLQMPGFTSIIQVEYDVLSDISHKYLRTQTV 274
QY 194 LDFMENFHOEHEHFKQEVSKELIGLVVLTKNKTYRVDDIDWDQNKSTFKKADGSE 253
DB 275 WEVWMDLFRKA-RGRFKEIETKMTGQIVLTKNKTYRVDDIDFDTPADTFETRSR-G-P 332
QY 254 VSFLEYKQKQNETLTKQPVTVSQPRRRGPGTLLGPAMLLPELCYLTGLTDKMRND 313
DB 333 VSYVDYFKKSYERVIHDVQNPWLVSRRPKREKGV--GPAYILPELCLFTGLSDMRAD 390
QY 314 FVWMDLAVHTRLTPEORQREVGRLIDYIHKNQNVQRELQWGLSFDSENLSFSGRILOT 373
```

```
DB 391 FVWMDKQGHTRVSPQDRCTLSGFIKKLSNEVEKYTLDSNGMEFDEKQYKLTGRVLP 450
QY 374 EKIHGGKTEPDYNFQFADMSKETEGAPLISVYPJDNMLIYTRNYESANSLQNLKRV 433
DB 451 EKLFGQKQFQSYNSNDMSDRTGNALTDKILNNWKIFYRRDANRGQDFIKSLVRA 510
QY 434 PAMGQMKALMIEY-DRTFAIYLVQOKYATDQTVVCLSSNRKQKDAIKKYCTD 492
DB 511 NPMGNVAGPEIIVLPDRITETYSLOAQIADQTVVLPTRKDRDAIKKTCVY 570
QY 493 CPTSPQCVARTLIGKQVTMAIAKRLALQNNCKMGELMRVDIPKLVIVGIDCYHMT 552
DB 571 HPCBQVIVSRTLSKQQLMSVATKIMQNNCKMGGLMRVEIPLSNMILGIDSYHSL 630
QY 553 AGRRSIAGFVASINEGTRMFSCRIFQDRGOELVDGLKVCLOALBANS CNEYMPRII 612
DB 631 TKGRSVLGFVASMNKSQTSFSSQAFQAGHGFANLSTLNNMLKRYQJNEKPEKII 690
QY 613 VYRDGVPDQGLKTLVNEVEPQLDCL-KSIRGYNPLTVIYVKKRVNTFFAOGGRLQ 671
DB 691 IFRDVGDSQYNLVVDYELKQIKDTLDKVTYQGTIVHKLAVVVKRIINRFFANLRGSL 750
QY 672 NPLPQTVIDEVTRPEWYDFIVSQAVSSVSTTHVNIYDNGSLKPHIORLTY 727
DB 751 NPPPGTVIDVVTKEHLYDYFIIISQVQGSVSPTSYVWVYDTGLKPDHMQRLTY 806
```

Search completed: May 5, 2004, 15:14:13
Job time : 55 secs